Genomics and transcriptional analysis of the Neodiprion abietis nucleopolyhedrovirus.

by

Simon Duffy
B.Sc., University of Victoria, 2001

A Dissertation Submitted in Partial Fulfillment
of the Requirements for the Degree of

DOCTOR OF PHILOSOPHY

in the Department of Biology

Genomics and transcriptional analysis of the Neodiprion abietis nucleopolyhedrovirus

by

Simon Duffy
B.Sc., University of Victoria, 2001

Supervisory Committee

Dr. D.B. Levin, (Department of Biology)

_____

Supervisor

Dr B. Koop, (Department of Biology)

_____

Departmental Member

Dr. N. Sherwood, (Department of Biology)

_____

Departmental Member

Dr. C. Upton, (Department of Biochemistry and Microbiology)

_____

Outside Member

Dr. D. Theilmann, (Agriculture and Agri-Food Canada, Summerland, BC)

_____

External Examiner

Supervisory Committee

Dr. David Levin
Supervisor

Dr. Ben Koop
Co-supervisor or Departmental Member

Dr. Chris Upton
Outside Member

Dr. David Theilmann
External Examiner

ABSTRACT

Baculoviruses are a family of mostly insect-specific viruses with relatively complex infection pathology. The baculovirus genome encodes between 89-181 genes, that are regulated by a complex temporal cascade of gene expression. Although baculoviruses pathogenic to lepidopteran hosts are well characterized, relatively little is known about non-lepidopteran baculoviruses. This thesis provides the genome sequence of a baculovirus pathogenic to the balsam fir sawfly (order: *Hymenoptera*), the *Neodiprion abietis* nucleopolyhedrovirus (NeabNPV). Our analyses of the NeabPNV genome indicated that the regulation of NeabNPV early genes by immediate early transactivators may differ from the lepidopteran baculovirus model. Also, through genome sequence analysis, we propose a model for the evolution of sawfly baculoviruses that is mediated by interspersed genome repeats. By selecting key genes transcribed at distinct time points during baculovirus infection as well as quantifying viral DNA in the tissues of

infected host larvae, we mapped the progression of NeabNPV infection *in vivo*. Based on the temporal scale of viral infection, we were able to show that two putative zinc-finger proteins, neab24 and neab52, are expressed in the immediate early and early stages of infection. The hypothesis that the mechanism of early gene regulation in non-lepidopteran baculovirus differs from that of lepidopteran baculoviruses prompted us to investigate promoter elements of 30 baculovirus genomes sequenced to date, *in silico*. This analysis revealed some sequence motifs may represent promoter elements in a wide range of baculoviruses, and that there may be differences in regulation of transcription between genes of the same temporal class in different baculovirus species.

Supervisor: Dr David Levin, (Department of Biology)

Table of Contents

List of Abbreviations

BV Budded Virus
DNA Deoxyribonucleic Acid
FPL Forest Protection Limited
GV Granulovirus
HMM Hidden Markov Model
hpi Hours Post Infection
NCBI National Center for Biotechnology Information
NPV Nucleopolyhedrovirus
NSERC Natural Sciences and Engineering Research Council
OB Occlusion Bodies
ODV Occlusion Derived Virion.
ORF Open Reading Frame
PCR Polymerase Chain Reaction
PIB Polyhedral Inclusion Body
PMRA Pest Management Regulatory Agency
PWM Position-Weighted Matrix
QPCR Quantitative PCR
RNA Ribonucleic Acid
RT-PCR Reverse Transcription PCR
SDS Sodium Dodecyl Sulfate
TBE Tris Borate EDTA buffer
TF Transcription Factor
TSS Translation Start Site
UV Ultraviolet
WSSV White Spot Shrimp Virus

List of Virus Species Abbreviations

AcMNPV *Autographa californica* NPV
AdhoNPV *Adoxophyes honmai* NPV
AdorGV *Adoxophyes orana* GV
AgseNPV *Agrotis segetum* GV
BmNPV *Bombyx mori* NPV
CfMNPV *Choristoneura fumiferana* NPV
CfMNPV-def *Choristoneura fumiferana* defective NPV
ChchNPV *Chrysodeixis chalcites* NPV
CpGV *Cydia pomonella* GV
CrleGV *Cryptophlebia leucotreta* GV
CuniNPV *Culex nigripalpus* NPV
EppoNPV *Epiphyas postvittana* NPV
HabrGV *Harrisina brillians* GV
HearNPV *Helicoverpa armigera* NPV
HearNPV-G4 *Helicoverpa armigera* NPV-G4
HzNPV *Helicoverpa zea* NPV
LdMNPV *Lymantria dispar* NPV
MacoNPV-A *Mamestra configurata* NPV-A
MacoNPV-B *Mamestra configurata* NPV-B
NeabNPV *Neodiprion abietis* NPV
NeleNPV *Neodiprion lecontei* NPV
NeseNPV *Neodiprion sertifer* NPV
OpMNPV *Orgyia pseudotsugata* NPV
PhopNPV *Phthorimaea operculella* GV
PxGV *Plutella xylostella* GV
RaouNPV *Rachiplusia ou* NPV
SpexNPV *Spodoptera exigua* NPV
SpliNPV *Spodoptera littoralis* NPV
SpltNPV *Spodoptera litura* NPV
TnSNPV *Trichoplusia ni* NPV
XcGV *Xestia c-nigrum* GV

List of Tables

List of Figures

Acknowledgments

I would like to thank my Supervisor, Dr. David Levin, for providing me an opportunity to pursue this research and for his constant optimism and support. As well, I thank the Levin lab members who have been such an important part of my life in the past five years. I particular, I would like to thank Aaron Young and Roberto Alberto for their mentorship. Thanks to Benoit Morin and Jen Beltgens for their assistance in screening genome libraries, and to Elisa Becker and Supipi Kaluarachchi for helping to optimize protocols for RT-PCR and QPCR. Also, I thank Beatrixe Whittome for providing tissue extracts of infected balsam fir sawfly larvae. I would also like to thank David Harrison and Tina Thorne for their support and advice over the past few years.

In addition, I would like to extend my gratitude to Dr. Ben Koop and the Koop lab. The Koop lab showed a tireless commitment to the genome sequencing effort. Without their help, understanding, and patience this project would have never succeeded.

I would like to thank Chris Lucarotti and the Canadian Forest Service (CFS) for partial funding and for coordinating the project. Similarly, I would like to thank NSERC and the Biocontrol Network of Canada for funding.

Dedication

I would like to dedicate this thesis to the memory of John Fraser. We became close friends during my graduate program and his passion and commitment to teaching and research was a true inspiration.

Chapter 1. General Introduction

1.1. Baculovirus

The *Baculoviridae* is a family of viruses that are pathogenic to arthropods. The baculovirus can be traced from the literature as early as the sixteenth century. It was described as a "wilting disease" infecting silk-producing larvae. A baculovirus was first commercialized as a viral insecticide in the 1960s when the *Helicoverpa zea* SNPV (HzSNPV) was registered under the name Virion H to control the cotton bollworm *H. zea* (Shieh, 1989). More recently, baculoviruses have been extensively used as vectors for expressing heterologous proteins cultured in insect cells (Jarvis, 1997; Shieh, 1989). Modified NPV vectors have also been used for efficient transient and stable transduction of mammalian cells (Condreay *et al.*, 1999).

1.2. Baculovirus Taxonomy

Most families of arthropod viruses also have members that infect vertebrates (Federici, 1997). The *Reoviridae* and *Rhabdoviridae* have the greatest pathogenic range with members that infect arthropods, vertebrates, and plants. The *Poxviridae*, *Iridoviridae*, *Parvoviridae*, *Reoviridae*, and *Picornaviridae* all have members that are pathogenic to arthropods and vertebrates. The *Baculoviridae* differ from the others by being the only family of viruses restricted to arthropods. The baculoviruses, poxviruses, and iridoviruses are the most closely related in that they all have large double-stranded DNA genomes. However, while the poxviruses and iridoviruses have linear DNA genomes, the baculoviruses have circular DNA genomes.

More than 600 baculoviruses have been reported in a multitude of arthropod hosts but the vast majority have been identified as pathogens for the members of the class *Insecta*, primarily in the orders *Hymenoptera*, *Diptera* and *Lepidoptera* (Blissard and Rohrmann, 1990). The taxonomy of baculoviruses has been resolved best for those that infect lepidopteran hosts. Traditionally, baculovirus names are derived from the species of the host insect followed by the viral genera. These lepidopteran baculoviruses have two genera, *nucleopolyhedroviridae* (NPV) and *granuloviridae* (GV). The type species for the NPVs is the *Autographa californica* multiple NPV (AcMNPV) while the type species for the GVs is the *Cydia pomonella* GV (CpGV). Baculovirus taxonomy generally corresponds to phylogenetic topology, as illustrated in Figure 1.1.

Lepidopteran NPVs and GVs are characterized based on the morphological properties and subcellular localization of the occluded virion. The lepidopteran NPVs are occluded in large (0.5μm –5μm diameter) protein crystals (Federici, 1997), called polyhedral inclusion bodies (PIBs), or simply occlusion bodies (OBs), that are synthesized in the nucleus of the host cell. Within the lepidopteran NPVs there are two distinct cladistic groups based on phylogenetic topology, the Group-I and Group-II NPVs (Zanotto *et al.*, 1993). Another characteristic used to distinguish lepidopteran NPVs involves whether the rod-shaped viral nucleocapsids are enveloped singly (SNPV) or multiply as a group (MNPV) within a membranous envelope, subsequently occluded within PIBs.

Lepidopteran GVs are similar in structure to NPVs except that they are singly enveloped and occluded within smaller structures, of 0.3μm × 0.5μm, referred to as granules (Federici, 1997). Whereas the NPV polyhedra are nuclear-localized, GV granules are typically observed in the cellular cytoplasm following nuclear disintegration.

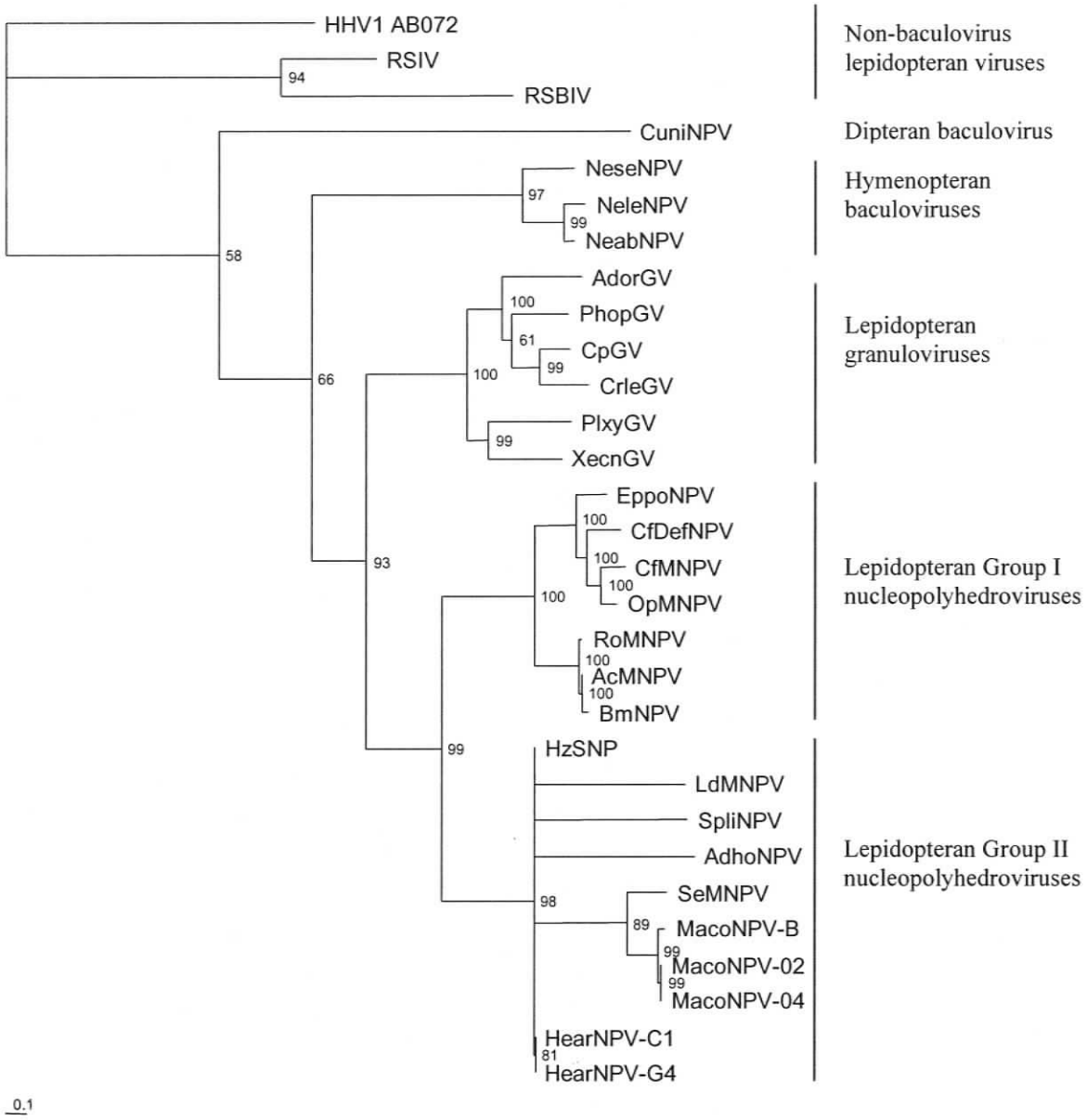


Figure 1.1. Phylogenetic analysis illustrating the distinct clades of the *Baculoviridae*. This analysis represents a modification of the phylogenetic analysis of DNA polymerase by Garcia-Maruniak *et al.* (2004). The data set was increased to include NeabNPV. Sequences were aligned using ClustalX 1.84 and a maximum parsimony tree was constucted (1000 bootstrap replicates) using PAUP 4.0 b10.

GV cladistics are defined by the different tissue trophisms that the viruses exhibit. Type 1 GVs invade the midgut epithelium in a similar manner as lepidopteran NPVs, but subsequently only infect fat body tissue, whereas the pathology of type 2 GV parallels that of the lepidopteran NPV (Federici, 1997). The third GV clade has only one member, *Harrisina brillians* GV (HabrGV). HabrGV gross pathology is restricted only to midgut epithelium, where it rapidly spreads from cell to cell (Federici and Stern, 1990). Death from HabrGV infection occurs at the same rate as typical lepidopteran NPVs, although host mortality is predominantly due to their inability to retain water within their body tissues.

To date, NPVs have only been isolated from insects in the orders *Hymenoptera* and *Diptera* (Federici, 1997). Recent genomic and morphological analyses indicate that the NPVs of the hymenopteran, dipteran and lepidopteran insects represent distinct taxa (Afonso *et al.*, 2001; Garcia-Maruniak *et al.*, 2004; Herniou *et al.*, 2004; Lauzon *et al.*, 2004; Moser *et al.*, 2001). As with the GVs, all non-lepidopteran NPVs identified to date have been singly enveloped (Federici, 1997). In contrast to lepidopteran baculoviruses, most non-lepidopteran NPVs appear to be restricted to midgut epithelium (Federici, 1997).

The NPV of the dipteran mosquito, *Culex nigripalpus*, has morphological and developmental characteristics that make it distinct from the hymenopteran and lepidopteran baculoviruses. The *C. nigripalpus* NPV (CuniNPV) is classified as a NPV despite the fact that, unlike other NPVs, CuniNPV OBs are typically smaller than the GV OBs, at 0.5μm in diameter (Moser *et al.*, 2001). As with the lepidopteran NPVs, however, the OBs are exclusively found in the nucleus (Moser *et al.*, 2001). On the other

hand, dipteran NPVs form two clades based on tissue trophism, which is similar to lepidopteran GVs. Type 1 dipteran NPVs are restricted to midgut epithelium while type 2 dipteran NPVs, comprised only of the *Tripula palusida* NPV, infect midgut epithelium and hemocytes (Federici, 1997).

Hymenopteran NPV OBs are smaller than those of lepidopteran NPVs (unpublished), at 0.5-1μm in diameter, and accumulate in the nucleus of the host cell. Neither the process of infection nor the pathology induced by hymenopteran NPVs has been well characterized. Recent genome sequence anlayses however, have begun to elucidate the structure and organization of hymenopteran NPVs as well as phylogenetic relationships between them (Garcia-Maruniak *et al.*, 2004; Herniou *et al.*, 2004; Lauzon *et al.*, 2004).

1.3. Baculovirus Infection Cycle

1.3.1. Infection

Baculovirus OBs can persist in contaminated soil and foliage for over ten years (Olofsson, 1988; Thompson and Scott, 1979). Transmission of a baculovirus occurs via ingestion of contaminated material and infection is initiated when the OB is dissolved under the highly alkaline conditions of the insect host midgut. The dissolution is aided by alkaline proteases that reside in the midgut lumen. The released virions pass through the peritrophic membrane and undergo receptor-mediated fusion with midgut epithelial cell plasma membrane. Within the cytoplasm of the target cell, nucleocapsids are released from the virion and migrate to the nucleus via actin cables of the cytoskeleton (Charlton and Volkman, 1993). Within the nucleus, the viral DNA is released from the nucleocapsid to initiate viral transcription and DNA replication (Blissard, 1996).

1.3.2. Early Gene Expression

Gene expression in lepidopteran baculoviruses has been well characterized and is divided into distinct stages including immediate-early, delayed-early, late, and very-late. Immediate-early genes are transcribed in a manner independent of other viral gene products (Guarino and Summers, 1986; Guarino and Summers, 1987). Transient early and late transcription and DNA replication assays suggest that four virus-encoded proteins, IE-0, IE-1, IE-2, and PE38, *trans*-regulate early gene transcription (Miller *et al.*, 1983; Ribeiro *et al.*, 1994; Wu *et al.*, 1993a; Wu *et al.*, 1993b). Early-expressed genes are transcribed by the host RNA polymerase II and contain typical eukaryotic consensus transcription motifs such as a TATA and CAGT motif. The primary result of early gene expression in baculoviruses is the replication of viral DNA. In contrast, hymenopteran and dipteran baculoviruses have been poorly characterized, and early gene expression appears to differ significantly from the lepidopteran baculoviruses, in that their genomes do not encode orthologues for any of the immediate early genes: *ie-0*, *ie-1*, *ie-2*, and *pe38* (Afonso *et al.*, 2001; Garcia-Maruniak *et al.*, 2004; Lauzon *et al.*, 2004).

1.3.3. Late Gene Expression

Lepidopteran baculovirus late genes are transcribed by an α-amanitin resistant viral-encoded RNA polymerase that initiates from the consensus A/GTAAG motif. This RNA polymerase is thought to be encoded by the genes *lef-6*, *lef-8*, and *lef-9* (Lu and Miller, 1994; Passarelli and Miller, 1994; Passarelli *et al.*, 1994). In addition to the genes involved in DNA replication, five additional genes (*lef-4*, *lef-5*, *lef-7*, *lef-11*, *p47*, and *39K*) are required for optimal expression from the late promoters (Todd *et al.*, 1995). The two genes, *ie-1* and *ie-2*, act as transactivators of both early and late baculovirus

gene expression. The major roles of baculovirus late and very late proteins are as structural elements of the virion.

1.3.4. Virion Structure

The lepidopteran baculoviruses have two distinct virion phenotypes; the budded virus (BV) and the occlusion derived virion (ODV). The nucleocapsid and BV have similar structures. The viral protein p6.9 binds the viral DNA (Wilson and Consigli, 1985) and is believed to be involved in DNA condensation because it is arginine-rich, like protamine condensation proteins. AcMNPV viral DNA is encapsulated by a nucleocapsid composed of vp39 (*ac89*), p87 (*ac104*), p24 (*ac129*) and pp78/83 (*ac9*) (Lu and Carstens, 1992; Miller *et al.*, 1990; Russell *et al.*, 1991; Wolgamot *et al.*, 1993). Of these, only vp39 gene orthologues have been identified in all baculovirus genomes sequenced to date (Garcia-Maruniak *et al.*, 2004; Lauzon *et al.*, 2004).

The BV is primarily responsible for cell-to-cell propagation of the virus. The BV is formed when the nucleocapsid interacts with the host cell plasma membrane where the viral protein GP64 accumulates (Volkman *et al.*, 1984). A mature GP64 is heavily glycosylated and acetylated (Roberts and Faulkner, 1989) and functions as a membrane fusion protein. The BV appears in electron micrographs as rod-shaped nucleocapsids loosely associated with a lipid-containing membrane. Peplomer structures are concentrated at one end of the virion and contain GP64. The GP64 protein is involved in membrane fusion, activated at low pH, which permits viral entry into the target cell cytoplasm following endocytosis (Blissard and Wenz, 1992; Monsma and Blissard, 1995).

In OBs, the nucleocapsids are singly or multiply embedded into a structure called the occlusion derived virion (ODV). Nucleocapsids destined for incorporation in an ODV appear to acquire a lipid envelope in the nucleoplasm from nuclear microvesicles (Braunagel *et al.*, 1996a). The structural region between the nucleocapids and the ODV envelope is called the tegument, which contains a protein called gp41 that is conserved in all baculovirus genomes sequenced to date (Garcia-Maruniak *et al.*, 2004; Lauzon *et al.*, 2004). One protein, ODV-EC27 (*ac144*), present in both the nucleocapsids and envelope of the ODV, is the only nucleocapsid protein present in the ODV but not the BV (Braunagel *et al.*, 1996b). Four other proteins are known to be associated with the ODV envelope: ODV-E66 (*ac46*), ODV-E25 (*ac94*), ODV-E56 (*ac148*), and ODV-E18/35 (*ac143/ac144*) (Braunagel *et al.*, 1996a; Braunagel *et al.*, 1996b; Hong and Summers, 1994; Russell and Rohrmann, 1993). The protein p74 (*ac138*) is embedded in the outermost surface of the ODV and is believed to be involved in receptor-mediated fusion with midgut epithelial cell plasma membrane during infection. ODV-EC27, ODV-E56, and p74 are conserved in all baculovirus genomes sequenced to date (Garcia-Maruniak *et al.*, 2004; Lauzon *et al.*, 2004).

Multiple individual ODV capsids are embedded in OBs. In lepidopteran baculoviruses, the OB is primarily composed of polyhedrin protein and the fibrous p10 protein. Polyhedrin is encoded in all sequenced lepidopteran and hymenopteran baculovirus genomes but not in the sequenced dipteran baculovirus (Afonso *et al.*, 2001). Polyhedrin and p10 are products of very late gene expression and rely on stimulation by the very late factor (VLF-1) (Todd *et al.*, 1996).

1.4. Baculovirus for Control of the Balsam Fir Sawfly

1.4.1. Baculoviruses as Insect Control Agents

Baculoviruses are natural insect-specific pathogens that are particularly attractive as bioinsecticides because they frequently cause epizootics in insect populations. Infection often results in death of the host by liquefaction, but sublethal infection may result in slower developmental rates, lower pupae and adult weights, higher long-term mortality, and reduced fecundity (Rothman and Myers, 1996). Baculoviruses are typically pathogenic to a narrow range of permissive insect species, but some have a broader host range, such as the alfalfa looper, *A. californica* NPV, that infects 35 species in the genus *Noctuidae*. Glaser and Chapman (Glaser and Chapman, 1913) first observed that the wilt disease of the gypsy moth (*Lymantria dispar*) was related to previously observed polyhedral diseases and that the agent responsible for this disease was a filterable virus. They also observed that the mode of virus infection was by oral ingestion and that the virus did not appear to be airborne (Glaser and Chapman, 1913). Steinhaus (1948) discussed the potential economic importance of a baculovirus pathogenic to the alfalfa butterfly, *Colias eurytheme*. He believed that the aggregate toll that the baculovirus placed on the alfalfa butterfly was one of the most important natural checks to control the insect population. The value of the disease as a control agent, however, was limited beacuase its effect is not exerted until after the crop is already significantly damaged (Steinhaus, 1948).

In the fifty years since these observations, interest in baculovirus-based biopesticides has both risen and waned. The early product Elcar (*Helicoverpa zea* NPV) was unable to compete with products based on synthetic pyrethroids in control of the cotton bollworm (Copping and Menn, 2000). In addition to Steinhaus' observation that baculoviruses are

effective only after the crop has been damaged (Steinhaus, 1948), there were additional limitations to using baculoviruses as biopesticides. Although some baculovirus occlusion bodies are stable in the soil for over 10 years and at 4°C for over 21 years, liberated baculovirus particles appeared to quickly lose their infectivity if exposed to sunlight ultraviolet (UV) between 250-280 nm (Smirnoff, 1972) and environmental conditions. Another limitation to spraying crops with baculovirus suspensions was that crops, such as cotton, might secrete alkaline dew (ca pH 8.2-9.1) (Andrews *et al.*, 1973) or highly alkaline or acidic soil (Thomas *et al.*, 1973) might liberate and inactivate occluded virions. Despite these early set-backs, however, improvements in the spray formulations derived from deliberately impure virus preparations combined with UV protectants have drastically improved the efficacy of baculovirus pesticides (David *et al.*, 1971; Jaques, 1971; Jaques, 1972).

The age of genomics allowed researchers to address Steinhaus' concern that baculoviruses were too slow in killing the insect host (Steinhaus, 1948). One approach involved modifying baculoviruses using recombinant technologies to over-express insect hormones resulting in abnormal growth, feeding, and/or death of the insect host. An effective approach involved modifying the NPV pathogenic to *Bombyx mori* (BmNPV) to express a functionally active diuretic hormone (BmDH5) and caused death in the host 20% faster than the wildtype virus (Maeda, 1989). The most successful approach involved introduction of the highly potent insect-selective toxin AaIT from the scorpion *Androctonus australis*. This increased the speed of kill by 40% (Maeda *et al.*, 1991; McCutchen *et al.*, 1991; Stewart *et al.*, 1991) and decreased feeding damage to cotton by 60% (Cory *et al.*, 1994; Hoover *et al.*, 1995).

In the United States, five NPVs and two GVs have been registered as biopesticides (source: http://www.epa.gov/pesticides/biopesticides/ingredients/index.htm). All of these products utilize lepidopteran baculoviruses as the active ingredient. In Canada, baculovirus pesticide products are registered for use against the lepidopteran Douglas fir tussock moth and gypsy moth, and for control of the hymenopteran red-headed pine sawfly *Neodiprion lecontei* (source: Canadian Pest Management Regulatory Agency ELSE database; http://eddenet.pmra-arla.gc.ca/4.0/4.0.asp). There has been a significant amount of research pioneering the use of sawfly-infecting baculoviruses to control insect populations in Canada.

The sawfly baculovirus, pathogenic for the European spruce sawfly (*Gilipinia hercyniae*) and the European pine sawfly (*Neodiprion sertifer*), was inadvertently introduced to North America and early investigation suggested that baculovirus introduction could result in significant and ongoing natural control (Balch and Bird, 1944 Bird and Elgee, 1957). Experimental ground applications of the sawfly NPV pathogenic to the balsam fir sawfly showed that this virus could be used to control outbreaks of this insect (Olofsson, 1973). This research supported the registration of the baculovirus control agent Neocheck-S that employed the natural NPV pathogen for *N. sertifer*.

1.4.2. The Balsam Fir Sawfly

The balsam fir sawfly (*Neodiprion abietis* Harris) (Hymenoptera: Diprionidae) is a major defoliator indigenous to North America. The larvae primarily feed on the foliage of balsam fir (*Abies balsamea* [L.] Mill) but are also commonly found on white (*Picea glauca* [Moench.] Voss) and black (*Picea mariana* [Mill.] B.S.P) spruce. The larvae prefer to feed on 1-2 year old needles and this defoliation results in increased tree

mortality and reduced tree growth. The balsam fir sawfly only directly damages the tree during the larval stage of its life cycle.

The adult female balsam fir sawfly has a 'saw-like' ovipositor with which she cuts a small slit in the edge of a needle to lay her egg. There is typically only one egg placed in each needle and they are usually laid in the current year's growth. The eggs overwinter during this stage and hatch as the tree buds open in the spring. The hatching larvae feed in colonies of 30-100 larvae, concentrating on the previous year's growth. The larvae strip the outside of the needle, leaving the central filament that in time reddens in colour and shrivels. The larvae go through 6 instars before they are mature and begin to pupate around mid-June.

There is only one generation of balsam fir sawfly every year and, when outbreaks occur, they typically last 3-4 years. The incidence and length of outbreaks is greater in pre-commerically thinned balsam fir stands. Recently, a number of outbreaks have occurred in managed stands of Atlantic Canada. Defoliation resulting from balsam fir sawfly is not uncommon in Atlantic Canada but the current outbreak is apparently unprecedented in severity (Piene *et al.*, 2001).

1.5. Thesis Outline

The goal of this thesis is to characterize the genome and gene expression of the baculovirus pathogenic to the balsam fir sawfly, the *Neodiprion abietis* nucleopolyhedrovirus (NeabNPV). The second chapter of the thesis reports the complete genome sequence of the NeabNPV as well as a comparative analysis of sawfly-infecting baculovirus genomes. Also, chapter two will examine the traditional approach towards characterization baculovirus gene promoter sequences and propose alternate additional analyses.

Chapter three examines the temporal patterns of DNA replication and gene expression of NeabNPV *in vivo*. Quantitative PCR (QPCR) was used to examine viral DNA kinetics and reverse transcription PCR (RT-PCR) of NeabNPV transcripts with sequence homology to characterized lepidopteran baculovirus genes was used to infer the timing of early, late and very late gene expression.

Chapter four reports the sequence and transcriptional analysis of two novel NeabNPV early expressed genes, which encode zinc finger motifs, a common feature of some baculovirus transactivators.

Chapter five presents a more refined analysis of baculovirus promoters. This analysis is an expansion of the analysis of predicted NeabNPV promoters presented in chapter two, but includes all 29 baculovirus genomes fully sequenced at the time of this analysis plus NeabNPV.

Chapter six summarizes the contributions of this thesis to the study of sawfly baculoviruses and discusses how these techniques can be applied to expand our understanding of non-lepidopteran baculoviruses.

Chapter 2. Genome sequence analysis of the *Neodiprion abietis* nucleopolyhedrovirus.

2.1. Abstract

Of thirty baculovirus genomes that have been sequenced to date, the only non-lepidopteran baculoviruses include the dipteran *Culex nigripalpus* nucleopolyhedrovirus (CuniNPV) and two hymenopteran nucleopolyhedroviruses that infect the sawflies, *Neodiprion lecontei* (NeleNPV) and *Neodiprion sertifer* (NeseNPV). This study provides a complete sequence and genome analysis of the nucleopolyhedrovirus that infects the balsam fir sawfly, *Neodiprion abietis* (Hymenoptera, Symphyta, Diprionidae). The *N. abietis* nucleopolyhedrovirus (NeabNPV) is 84,264 bp in size, with a G+C content of 33.5%, and contains 93 predicted open reading frames (ORFs). Eleven predicted ORFs are unique to this baculovirus, 10 ORFs have a putative orthologue in the NeleNPV genome but not the NeseNPV genome, and one ORF (*neab53*) has a putative orthologue in the NeseNPV genome, but not the NeleNPV genome. Specific repeat sequences are coincident with major genome rearrangements that distinguish NeabNPV and NeleNPV. Genes within these repeat regions encode a common amino acid motif, suggesting that they are a family of repeated contiguous gene clusters. Lepidopteran baculoviruses, similarly, have a family of repeated genes called the *bro* gene family. However, there is no significant sequence similarity between the NeabNPV and *bro* genes. Orthologues of early-expressed genes such as *ie-1* and *lef-3* were absent in NeabNPV, as they are in the previously sequenced hymenopteran baculoviruses. Analyses of ORF upstream sequences identified potential temporally distinct genes based on putative promoter elements.

2.2. Introduction

Balsam fir sawfly (*Neodiprion abietis,* Hymenoptera, Symphyta, Diprionidae) is a native sawfly species that occurs throughout Canada and the United States. The balsam fir sawfly is a defoliating species that feeds primarily on balsam fir, *Abies balsamea* (L.) Mill. and spruce, *Picea* spp. (Wallace and Cunningham, 1995). An outbreak of the balsam fir sawfly in Newfoundland has persisted since 1991. This 15-year outbreak is in contrast to the typically observed outbreaks of 4-5 year durations that occur every 5-15 years. The current infestation, spanning over 40,000 ha of balsam fir in western Newfoundland, threatens the significant silivicultural investment in that region, as severe defoliation greatly reduces tree growth and may cause tree mortality (Piene *et al.*, 2001). The application of a virus naturally pathogenic to the balsam fir sawfly, the *Neodiprion abietis* nucleopolyhedrovirus (NeabNPV), may offer great potential in controlling sawfly populations (Olofsson, 1973). Recently, aerial application of NeabNPV have successfully initiated collapse of growing and peaking balsam fir sawfly populations (Moreau *et al.*, 2005).

NeabNPV is a member of the *Baculoviridae*, a well-studied family of invertebrate viruses that have large double-stranded circular DNA genomes within rod-shaped nucleocapsids that are enveloped and occluded. The two genera of baculoviruses, *Nucleopolyhedrovirus* (NPV) and *Granulovirus* (GV) are typically distinguished by the size and localization of their occlusion bodies (OBs). NPVs have larger nuclear-localized OBs, ranging in diameter from 1-5 μm, with singly or multiply enveloped virions. GVs have smaller OBs, with a diameter between 0.3-0.6 μm, in which single, enveloped virions are normally occluded and these OBs are distributed throughout the cell following

nuclear disintegration. GVs have only been isolated from lepidopteran hosts, while NPVs have been isolated from Lepidoptera, Hymenoptera and Diptera. The *Neodiprion abietis* nucleopolyhedrovirus (NeabNPV) is a singly enveloped NPV that infects the balsam fir sawfly.

To date, 30 baculovirus genomes are available in GenBank. Of these, the lepidopteran baculoviruses are most highly represented (25 sequenced genomes). Two NPVs that infect diprionid sawflies, *N. sertifer* (NeseNPV) (Garcia-Maruniak *et al.*, 2004) and *N. lecontei* (NeleNPV) (Lauzon *et al.*, 2004), and one that infects the mosquito, *Culex nigripalpus* (Diptera, Culicidae) (Afonso *et al.*, 2001) have been completely sequenced. Baculovirus genomes range in size from 81.8 Kb to 161 Kb and have G+C content ranging from 32% to 56%. The sawfly NPVs have both the smallest genomes and the lowest G+C content of any baculoviruses sequenced to date (Garcia-Maruniak *et al.*, 2004; Lauzon *et al.*, 2004).

Baculovirus phylogeny was first resolved using the conserved polyhedrin gene sequence. This analysis identified two distinct taxa within the lepidopteran NPVs, termed groups-I and -II (Zanotto *et al.*, 1993). More recently full genome phylogenetic analyses suggested that single gene phylogenies might not accurately resolve evolutionary relationships among baculoviruses (Herniou *et al.*, 2001). Instead, seven genes are considered to share the topology of whole genome phylogenies (Herniou *et al.*, 2001). Complete genome phylogeny, along with gene content and gene order phylogenies suggest that the Lepidoptera-infecting baculoviruses form three distinct clades; GV, Group I-NPV, and Group II-NPV (Herniou *et al.*, 2001; Herniou *et al.*, 2003). The Diptera-infecting baculovirus, CuniNPV, is taxonomically distinct from the

Lepidoptera-infecting baculoviruses (Herniou *et al.*, 2001). Recent genome sequencing of the Hymenoptera-infecting baculoviruses, NeseNPV and NeleNPV, has indicated that they represent a separate clade that is distinct from both Lepidoptera-infecting and Diptera-infecting baculoviruses (Garcia-Maruniak *et al.*, 2004; Herniou *et al.*, 2004; Lauzon *et al.*, 2004).

In this study, we report the genome sequence for NeabNPV. Our data provide further insight into the evolution of sawfly baculoviruses. Furthermore, a detailed analysis of the upstream sequences of predicted NeabNPV ORFs identifies promoter motifs that potentially discriminate temporally distinct genes. Twenty-nine genes are believed to be conserved throughout all fully sequenced baculovirus species. This set of genes, however, does not include a number of genes that are highly conserved among lepidopteran baculoviruses, such as, the early-expressed genes necessary for viral DNA replication (*ie-1* and *lef-3*) and gene regulation (*me53*, *lef-6*, and *pp31*). Due to the essential role that these genes play in the lepidopteran baculovirus life cycle, we investigated promoter characteristics that distinguish early- and late-expressed genes. The subset of predicted early-expressed genes might include functional analogues for these genes.

2.3. Materials and Methods

Source of NeabNPV. NeabNPV occlusion bodies (OBs) were produced and purified as described by Moreau *et al.* (Moreau *et al.*, 2005) This procedure involved thawing frozen (-20°C) larvae in five volumes 0.3% sodium dodecyl sulfate (SDS) solution followed by homogenization. The sample was twice treated with a process of filtration through a 1 mm^2 plastic mesh, followed by resuspension of solid matter in 0.3% SDS, and re-homogenization and re-filtration until the filtrate was clear. Filtrates were then pooled, filtered twice through eight layers of cheesecloth, and centrifuged for 15 min at 9,000 × g. The supernatant was discarded and the NeabNPV occluded body (OB) pellet was re-suspended in 0.3% SDS. Centrifugation and re-suspension were repeated until a clear supernatant was obtained. The NeabNPV OB pellet was then re-suspended in water and concentrated to 10^9 PIBs/mL.

Virus DNA purification. NeabNPV OB suspensions were washed with sterile ddH_2O and centrifuged (15,000 × g, 5 min at 20°C) three times and then treated with SDS (0.4%) under agitation for one hour. The resultant suspension was washed and centrifuged (15,000 × g, 10 min at 20°C) three more times and incubated with dissociation buffer (0.1M Na_2CO_3; 0.04M sodium thioglycolate; pH 10.8) for 20 min. Debris was pelleted twice by centrifugation (2,000 × g, 5 min at 4˚C) and the supernatant was collected. The supernatant was washed, pelleted (15,000 × g, 30 min at 4˚C) and resuspended in Tris-HCl. This was then incubated with 0.01 mg/mL proteinase K (Invitrogen) and 2% sarcosyl for 16 hours at 50˚C. NeabNPV DNA was extracted with DNAzol (Invitrogen) following the manufacturer's standard protocol and resolved by

0.8% agarose pulse-field gel electrophoresis in 0.5× TBE buffer (45 mM Tris-Borate; 1 mM EDTA; pH 8.0) at 200 V with a switch time of 60 s for 15 h followed by a switch time of 90 s for 8 h at 14°C. NeabNPV DNA was stained with 0.5 μg/mL ethidium bromide stain, viewed under long-wavelength UV illumination (312 nm) and excised. Excised DNA was purified using the Fermentas DNA Extraction kit (Fermentas) according to the manufacturer's protocol.

DNA cloning and Sequencing. NeabNPV DNA was fragmented by restriction enzyme digestion and by hydrodynamic shearing (HydroShear, GenMachine, San Carlos, CA). Partial restriction libraries were generated by digesting NeabNPV DNA with HindIII (New England Biolabs) and ligation into pBluescriptII KS+ (Stratagene) and digestion with EcoRI (New England Biolabs) followed by ligation into pT7/T3a18. Random libraries fragmented by hydrodynamic shearing, were generated and ligated into pSMART (Lucigen Corporation). PCR was used to generate gap-spanning fragments. The result was an average of 15-fold sequence redundancy. DNA sequencing was conducted using an ABI3700 automated DNA sequencer (Applied Biosystems, Inc.) and the reactions were carried out using ABI Prism BigDye Terminator Cycle Sequencing Ready Reaction Kit (Perkin-Elmer Applied Biosystems).

Genome Sequence Analysis. The complete NeabNPV genome was automatically assembled using PHRED/PHRAP/CONSED with a PHRAP repeat stringency of 0.8 (Ewing and Green, 1998; Ewing *et al.*, 1998). The contiguous sequence was manually edited and confirmed. Open Reading Frames (ORFs) were identified using NCBI ORF

Finder program (http://ncbi.nih.nlm.gov/gorf/gorf.html) and those encoding more than 50 amino acids with minimal overlap were considered as putative genes. Characterization of the ORFs involved BLAST analysis using BLASTP and PSI-BLAST against the NCBI non-redundant protein database (Altschul *et al.*, 1990), RPS-BLAST against the cdd database (Marchler-Bauer *et al.*, 2005). Baculovirus putative homologues were accepted based on a BLASTP (default parameters) E-value less than or equal to 0.001. Predicted ORFs were then analyzed with InterProScan (Zdobnov and Apweiler, 2001) that integrates a queries against PROSITE (Hofmann *et al.*, 1999), PRINTS (Attwood *et al.*, 2000), Pfam (Bateman *et al.*, 2000), ProDom (Corpet *et al.*, 1999), SMART (Schultz *et al.*, 2000), TIGRFAMs (Haft *et al.*, 2001), PIR superfamily (Wu *et al.*, 2003), SUPERFAMILY (Gough *et al.*, 2001), Gene3D (Buchan *et al.*, 2002) and PANTHER (Mi *et al.*, 2005) databases.

Genome Parity and Repeat Analysis. The repeat regions were identified using REPuter (Kurtz *et al.*, 2001). The 5' and 3' limits were defined as the outermost repeat with no other repeats within 1 Kb on the external side. The minimal repeat element was queried against the NeleNPV and NeseNPV genomes using BLASTN 2.2.6. Repeat regions were also illustrated by generating parity plots of the NeabNPV genome sequence against itself using PipMaker (Schwartz *et al.*, 2000). Additionally, genome parity between NeseNPV and NeabNPV, and NeleNPV and NeabNPV was illustrated using PipMaker (Schwartz *et al.*, 2000).

Promoter Sequence Analysis. Sequences, 160 bp upstream of each predicted ORF start codon, were analyzed for potential promoter motifs. These sequences were compared to experimentally verified promoter elements using SIGNAL SCAN v4.05 (Possee and Howard, 1987) to query the TRANSFAC database (Wingender, 1988). Potential promoter elements were identified by aligning all upstream sequences using AlignACE v3.0 (Hughes *et al.*, 2000). The sequences were also analyzed for common known baculovirus promoter elements including the TATA sequence (TATA, TATAW), CAKT, and DTAAG elements. These were also sought in the –40 bp to –20 bp region of the upstream sequences, relative to the predicted translation start site. Only putative predicted promoter elements with a p-value less than 0.05 were reported, exception for TATA elements that were over-represented in the –160 bp or –40 bp to –20 bp upstream sequences and a putative CAKT element, due to its established function as a baculovirus promoter element (Pullen and Friesen, 1995a).

Calculation of p-value. To generate a p-value for a sequence of a given length, 1000 random loci of equal length (160 or 20 bp) were sampled. The expected frequency of occurrence was calculated for each motif in this data set. The p-value for observing x motif-containing putative promoter elements out of N ORFs given an expected frequency F_e and a cumulative binomial distribution was calculated as follows:

$$P(\text{occ} \leq x) = \frac{N!}{x! \times (N-x)!} \times (F_e)^x \times (1- F_e)^{N-x}$$

Nucleotide Sequence and accession number. The NeabNPV genome sequence has been deposited in GenBank under accession no. DQ317692.

2.4. Results and Discussion

Nucleotide sequence analysis. The NeabNPV genome was 84,264 bp in size, making it among the smallest baculovirus genomes along with NeleNPV at 81,755 bp (Lauzon *et al.*, 2004), NeseNPV at 86,462 bp (Garcia-Maruniak *et al.*, 2004) and *Adoxophyes orana* GV (AdorGV) at 99,657 bp (Wormleaton *et al.*, 2003). The G+C content was equivalent to NeleNPV and NeseNPV at 33.5%. By comparison, the G+C content of lepidopteran baculovirus genomes ranges from 32.4% for *Cryptophlebia leucotreta* GV (CrleGV) (Lange and Jehle, 2003) to 57.5% for *Xestia c-nigrum* GV (XcGV) (Hayakawa *et al.*, 1999).

The NeabNPV genome contained 93 potential methionine-initiated ORFs (Figure 2.1, Table 2.1). This is greater than the number of ORFs predicted to be encoded by the NeleNPV (89 ORFs) or NeseNPV (90 ORFs) genomes (Garcia-Maruniak *et al.*, 2004; Lauzon *et al.*, 2004). Similar to the genome of NeseNPV, the NeabNPV ORFs are biased in orientation such that 60.9% are oriented clockwise and 39.1% are oriented in the opposite direction.

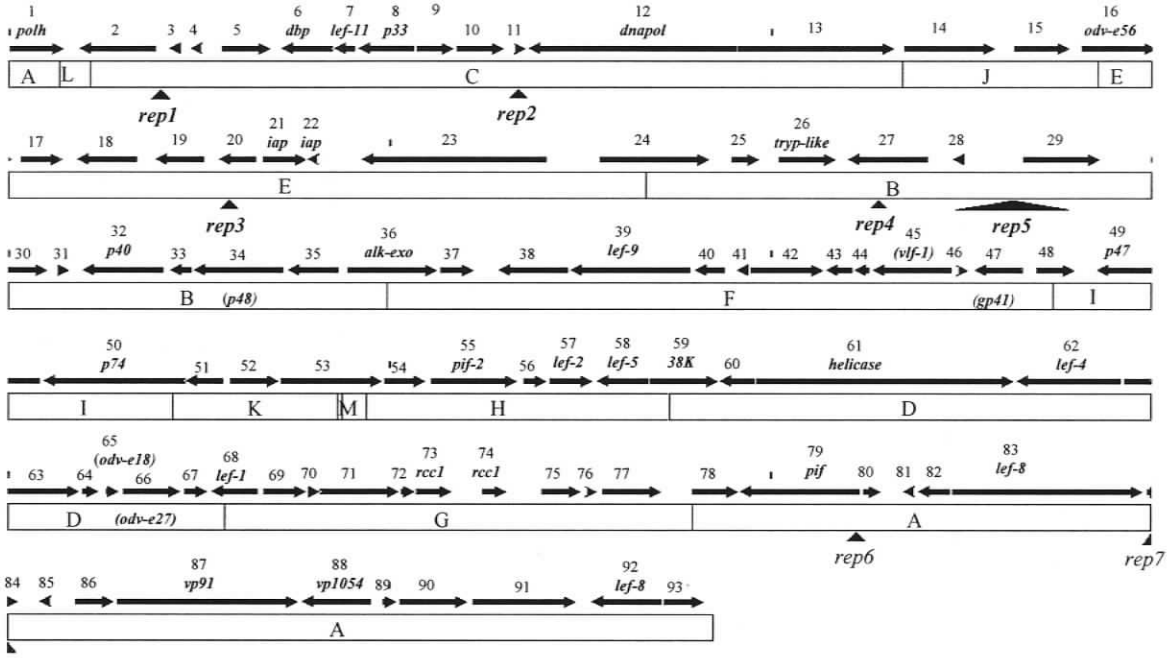


Figure 2.1. Linear sequence map and HindIII physical map of the circular NeabNPV genome. The transcriptional direction of each ORF is represented as an arrow labeled with the ORF number. Boxes beneath the arrows are labelled as the corresponding HindIII fragment. Known baculovirus predicted homologues are presented below the ORF number. Repeat regions (*rep*) are shown as triangles.

Table 2.1. NeabNPV ORFs

ORF	Name	Position	Length (aa)	Predicted Sequence Homologues					Feature of Coding Sequence[c]	Putative Promoter Elements[b]
				NeleNPV[a]	NeseNPV[a]	AcMNPV[a]	PxGV[a]	CuniNPV[a]		
1	*polyhedrin (polh)*	0>741	247	1 (99) 247	1 (93) 246	8 (46) 245	4 (36) 248			TATA Box; DTAAG; M7; M8; M9
2		906<1953	349	5 (84) 52	3 (14) 388				Cyclin-Like F-box Domain. RNA binding motif.	M1; M2
	rep1	1881>2755								
3		2093<2273	60							M1; M2
4		2380<2551	57							TATA Box; M1; M2; M4
5		2797>3469	224	13 (62) 247	20 (21) 270				Similar to AcMNPV P33. Signal peptide.	TATA Box; Zeste; DTAAG; M1; M6; M7; M8; M9
6	*dbp*	3557<4268	237	14 (88) 236	22 (45) 251	25 (19) 316				TATA Box; M2; M5
7	*lef-11*	4248<4557	103	15 (84) 101	23 (43) 100					TATA Box; CAKT; M9
8	*p33*	4570<5341	257	16 (87) 256	24 (51) 252	92 (19) 259	76 (18) 250	14 (14) 370		DTAAG; M3; M6
9		5342>5855	171	17 (91) 170	25 (47) 170					TATA Box; DTAAG
10		5872>6523	217	18 (77) 216	26 (44) 218				1 predicted TM domain. Signal peptide.	TATA Box; DTAAG; M7
	rep2	6570>6954								
11		6635>6809	58						2 predicted TM domains.	TATA Box; M3
12	*dnapol*	6795<9555	920	20 (84) 923	28 (64) 913	65 (23) 984	94 (23) 979	91 (17) 1138	DNA polymerase type-B family.	Zeste; CREB/ATF-47; CAKT; M6
13		9553>11638	695	21 (66) 728	29 (28) 792		93 (9) 651	67 (13) 1357	Putative ATPase. Putative IMP dehydrogenase domain. Prefoldin structure.	CREB/ATF-47; M3; M6; M7
14		11742>12963	407	22 (63) 380	3 (21) 388					
15		13175>13937	254	25 (35) 261					Leucine-rich repeat.	TATA Box; DTAAG; M3
16	*odv-e56*	14060>15071	337	23 (92) 336	38 (71) 339	148 (34) 376	16 (38) 351	101 (20) 361	1 predicted TM domain. IMP dehydrogenase / GMP reductase domain.	TATA Box; DTAAG; M5; M7; M9
17		15159>15738	193	24 (92) 192	39 (50) 192					TATA Box; DTAAG; M3
18		15870<16698	276	25 (52) 261						M3; M4; M9
19		16912<17596	228	26 (67) 226						TATA Box; Zeste; M1; M6; M9
20		17757<18267	170	2 (34) 108						TATA Box
	rep3	17804>18262								
21	*iap*	18332>18941	203	11 (75) 260	17 (24) 181	27 (23) 286	98 (21) 281		Baculovirus IAP Repeat Domain.	TATA Box; Zeste; M1; M7
22		18912<19074	54							CAKT; M7

23		19610<22079	823	9 (60) 794	16 (27) 637					TATA Box; M3
24		22749>24213	488	8 (50) 526	11 (16) 461				Ring finger domain. Zinc finger. LMP repeat.	TATA Box; M3
25		24466>24868	134	7 (71) 131						TATA Box; M3; M6; M7
26	*trypsin-like protein*	25091>25871	260	6 (96) 259	7 (73) 258				Trypsin-like serine protease.	DTAAG; M3; M6
27		25989<27069	360	4 (73) 332	9 (20) 416					TATA Box; Zeste; M4
	rep4	26312>26527								
28		27373<27559	62						Signal peptide.	Zeste; CAKT; M1; M6; M9
	rep5	27427>29532								
29		28291>29335	348	5 (84) 52	3 (14) 388				F-box domain.	TATA Box; CREB/ATF-47
30		29972>30539	189	27 (51) 190	37 (29) 133					TATA Box; M1; M2
31		30612<30869	86							TATA Box; CAKT; M5
32	*p40*	30950<32051	367	29 (91) 366	35 (64) 368					TATA Box; Zeste
33		32105<32423	106	30 (75) 116	34 (47) 117					TATA Box; DTAAG; M7
34	*p48*	32419<33628	403	31 (85) 390	33 (54) 402	103 (10) 387			IMP dehydrogenase / GMP reductase domain.	TATA Box; M6
35		33646<34348	234	32 (88) 238	32 (63) 235		40 (17) 206			Zeste; M6
36	*alk-exo*	34451>35663	404	33 (82) 402	31 (48) 399	133 (22) 419	106 (25) 378	54 (15) 367	IAP repeat. Restriction endonuclease-like.	
37		35659>36133	158	34 (84) 157		152 (14) 408			1 predicted TM domain. Signal peptide.	TATA Box; CAKT; DTAAG; M6; M9
38		36417<37371	318	36 (60) 319					1 predicted TM domain. Signal peptide.	TATA Box; M6
39	*lef-9*	37363<38953	530	37 (87) 503	40 (62) 507	62 (29) 516	99 (31) 494	59 (15) 590		TATA Box; M3; M7; M9
40		38981<39404	141	38 (87) 141	41 (48) 147	68 (20) 192	96 (15) 128	58 (18) 137	1 predicted TM domain.	TATA Box; M7
41		39533<39722	63							TATA Box; DTAAG; M7; M8; M9
42		39720>40707	329	39 (70) 353	42 (24) 444					Zeste; M3; M7
43		40703<41081	126	40 (87) 125	43 (39) 125					TATA Box; DTAAG; M7
44		41077<41308	77	41 (90) 76	44 (63) 82				1 predicted TM domain. Signal peptide.	CREB/ATF-47; M6
45	*vlf-1*	41310<42375	355	42 (86) 354	45 (65) 353	77 (24) 379	89 (29) 346	18 (16) 358	DNA breaking-rejoining enzyme, catalytic core. IMP dehydrogenase / GMP reductase domain.	TATA Box; CAKT, DTAAG

46		42458>42617	53						1 predicted TM domain. Signal peptide.	TATA Box
47	*gp41*	42649<43315	222	44 (94) 270	47 (60) 312	80 (30) 409	87 (19) 283			TATA Box; DTAAG; M4; M8
48		43484>44012	176	45 (90) 175	48 (75) 176	81 (36) 233	86 (34) 191	106 (14) 186	2 predicted TM domains.	Zeste; DTAAG; M3
49	*p47*	44255<45434	393	46 (87) 389	49 (61) 386	40 (22) 401	51 (20) 386	73 (10) 439		Zeste; CAKT; M7
50	*p74*	45435<47340	635	47 (86) 633	50 (63) 634	138 (38) 645	49 (31) 578	74 (32) 681	3 predicted TM domains. IMP dehydrogenase / GMP reductase domain.	TATA Box; M3
51		47317<47854	179	48 (69) 178	51 (28) 142				dsRNA-binding motif.	TATA Box; M8; M9
52		47916>48585	223	49 (76) 218	52 (40) 220				Zinc finger.	TATA Box; Zeste
53		48577>49933	452		53 (48) 386				1 predicted TM domain. IMP dehydrogenase / GMP reductase domain.	TATA Box; M4; M5; M6
54		49929>50508	193	51 (83) 192	54 (47) 184					TATA Box; CREB/ATF-47; DTAAG; M6; M9
55	*pif-2*	50537>51692	385	52 (93) 384	55 (73) 384	22 (43) 382	37 (46) 368	38 (43) 403	1 predicted TM domain. Signal peptide.	TATA Box; DTAAG; M3; M6; M7
56		51761>52094	111	53 (88) 104	56 (57) 115					TATA Box; DTAAG; M7
57	*lef-2*	52095>52683	196	54 (80) 195	57 (37) 200	6 (16) 210		25 (11) 225		TATA Box; M4
58	*lef-5*	52709<53417	236	55 (88) 235	58 (57) 230	99 (27) 265	69 (31) 247		Zinc finger.	TATA Box; M9
59	*38K*	53407>54325	306	56 (88) 305	59 (59) 304	98 (31) 320	70 (31) 340	87 (25) 303	IMP dehydrogenase / GMP reductase domain. Putative phosphatase.	Zeste; DTAAG
60		54310<54817	169	57 (92) 168	60 (63) 170	96 (25) 173	71 (24) 161	90 (25) 202	1 predicted TM domain. Signal peptide.	TATA Box; M5
61	*helicase*	54803>58214	1137	58 (89) 1134	61 (59) 1143	95 (18) 1221	72 (18) 1124	89 (12) 1332	IMP dehydrogenase / GMP reductase domain. P-loop containing nucleoside triphosphate hydrolases.	M7
62	*lef-4*	58210<59617	469	59 (85) 468	62 (52) 477	90 (22) 464	78 (25) 432	96 (16) 497		TATA Box; M4; M6
63		59633>60962	443	60 (89) 442	63 (52) 441	142 (20) 477				TATA Box; DTAAG; M3; M6
64		60970>61210	80	61 (89) 78	64 (43) 78				1 predicted TM domain	TATA Box; DTAAG; M4; M5
65	*odv-e18*	61285>61489	68	62 (95) 85	65 (83) 83				1 predicted TM domain. Signal peptide. Aspartate and glutamate racemases signature 1.	TATA Box; DTAAG
66	*odv-e27*	61503>62292	263	63 (94) 262	66 (71) 260	144 (21) 290	80 (12) 287			CAKT; DTAAG
67		62319>62652	111	64 (87) 110	67 (70) 108	145 (20) 77	12 (27) 98		1 predicted TM domain. Signal peptide. Putative chitin-binding motif.	TATA Box; DTAAG
68	*lef-1*	62659<63295	212	65 (85) 211	68 (54) 190	14 (20) 266	55 (27) 251	45 (26) 235	DNA primase.	TATA Box; CREB/ATF-47; M5

69		63352>63940	196	66 (89) 193	69 (63) 190	115 (24) 204	29 (28) 181	46 (27) 203	1 predicted TM domain. Signal peptide.	TATA Box; Zeste; M9
70		63939>64119	60						1 predicted TM domain. Signal peptide.	TATA Box; Zeste; DTAAG; M6; M7
71		64093>65158	355	67 (86) 354	70 (57) 357	109 (21) 390	43 (20) 414	69 (17) 409		TATA Box; M3; M7; M9
72		65157>65367	70	68 (89) 69	71 (53) 71				1 predicted TM domain.	TATA Box; DTAAG
73	*rcc1*	65353>65851	166	69 (80) 136	72 (60) 118				RCC1 domain. 1 predicted TM domain.	TATA Box; M3; M5; M6
74	*rcc1*	66214>66565	117	71 (71) 99	72 (7) 118				RCC1 domain. 1 predicted TM domain. Signal peptide.	TATA Box; M7
75		66996>67545	183	72 (64) 283	75 (39) 273					TATA Box; M7
76		67599>67755	52							TATA Box
77		67787>68585	266	74 (81) 265	77 (36) 284					TATA Box; M4; M7
78		68962>69592	210	75 (79) 209	78 (27) 143					TATA Box; CREB/ATF-47; M9
79	*pif*	69573<71187	538	76 (91) 530	79 (63) 519	119 (29) 530	7 (28) 536	29 (30) 523	1 predicted TM domain Signal peptide. Zinc finger.	Zeste; CREB/ATF-47; DTAAG; M1; M6; M9
	rep6	71014>72092								
80		71218>71485	89	35 (25) 97						TATA Box; CAKT; M7
81		71728<71899	57	2 (51) 108						TATA Box; M3; M6
82		71923<72379	152	77 (70) 149	80 (45) 156			28 (13) 284		TATA Box; CREB/ATF-47
83	*lef-8*	72377>74909	844	78 (87) 843	81 (69) 846	50 (30) 876	109 (32) 838	26 (20) 922	Subunits of DNA dependent RNA-polymerase.	TATA Box; Zeste; CAKT; M5
84		74926>75151	75	2 (28) 108						TATA Box
	rep7	74887>75140								
85		75391<75556	55							TATA Box; DTAAG; M4; M8
86		75878>76430	184	81 (89) 148	83 (66) 145				Capsid structural protein.	TATA Box; M1; M5; M8
87	*vp91*	76438>78844	802	82 (82) 803	84 (51) 824	83 (23) 847	84 (21) 533	35 (22) 741	1 predicted TM domain. Signal peptide. IMP dehydrogenase / GMP reductase domain. Chitin binding Peritrophin-A domain.	TATA Box; Zeste; DTAAG; M4; M7; M9
88	*vp1054*	78845<79787	314	83 (93) 313	85 (60) 310	54 (20) 365	115 (16) 311	8 (13) 329	IMP dehydrogenase / GMP reductase domain. DNA polymerase beta, N-terminal-like.	
89		79912>80125	71	84 (81) 69	86 (58) 73					
90		80132>81056	308	85 (83) 305	87 (38) 311					TATA Box; DTAAG; M6

91		81100>82480	460	87 (54) 476	18 (11) 301					TATA Box; M7
92	*vp39*	82640<83588	316	88 (98) 315	89 (71) 312	89 (18) 347	79 (14) 320	24 (10) 289	Putative RNA 2'-phosphotransferase	CAKT; M2; M5; M6
93		83597>84140	181	89 (80) 183	90 (43) 192				Putative RNA 2'-phosphotransferase	TATA Box; Zeste; DTAAG

[a] The ORF number of a putative homologue with the % amino acid identity in brackets followed by the length of the homologue.

[b] Promoter elements and abbreviations are defined in Table 2; TATA box was defined as a TATA sequence within the -40 to -20 upstream region of the ORF or a TATAW sequence within 160 bp upstream of the ORF; CAKT was reported only when it occurred within the -40 to -20 upstream region of the ORF; all other elements were reported if they occurred within 160 bp upstream of the ORF.

[c] Summary of BLAST, PSI-BLAST, RPS-BLAST, and INTERPRO SignalScan analyses.

NeabNPV Gene Content. Twenty-nine genes are shared among all baculovirus genomes sequenced to date (Garcia-Maruniak *et al.*, 2004; Lauzon *et al.*, 2004). Putative orthologues of all 29 of these conserved genes are found in NeabNPV. Of the 93 predicted NeabNPV ORFs, 72 have corresponding orthologues in NeseNPV and 81 have corresponding orthologues in NeleNPV. Eleven NeabNPV predicted ORFs are unique to NeabNPV (*neab3*, *neab4*, *neab11*, *neab22*, *neab28*, *neab31*, *neab41*, *neab46*, *neab70*, *neab76*, and *neab85*). Ten NeabNPV predicted ORFs (*neab15*, *neab18*, *neab19*, *neab20*, *neab25*, *neab37*, *neab38*, *neab80*, *neab81*, and *neab84*) have a putative sequence orthologue in the NeleNPV genome but not the NeseNPV genome, while one NeabNPV predicted ORF (*neab53*) has a putative sequence orthologue in the NeseNPV genome but not the NeleNPV genome.

Potential Envelope Fusion Proteins in Sawfly Baculoviruses. In AcMNPV, GP64 protein appears to be a critical protein for cell-to-cell transmission of the baculovirus budded virus, as *gp64*-deletion mutants were localized to insect midguts (Blissard *et al.*, 1992). In the *Lymantria dispar* NPV (LdMNPV) genome, where a *gp64* homologue was not identified (Kuzio *et al.*, 1999), the gene encoding a potential *gp64* analog, *ld130*, was proposed and its *Spodptera exigua* NPV (SpexNPV) sequence orthologue was later shown to mediate pH dependant membrane fusion between cells (IJkel *et al.*, 2000).

An orthologue for gp64 was not identified in NeabNPV, NeleNPV (Lauzon *et al.*, 2004), or NeseNPV (Garcia-Maruniak *et al.*, 2004) by BLAST analysis. Sequence similarity searches, however, may not be the most effective means of identifying analogs for envelope fusion proteins, as LdMNPV Ld130 bears no significant sequence similarity

to AcMNPV GP64. Based on the original criteria that proposed Ld130 as a potential envelope fusion protein (Kuzio *et al.*, 1999), candidate protein analogs should possess both a N-terminal signal and transmembrane domain.

Two NeabNPV ORFs, *neab10* and *neab44*, possess these characteristics and may be candidate envelope fusion genes. InterProScan analysis (see Materials and Methods section) of the predicted gene products indicated the presence of both a N-terminal signal and transmembrane domain. The putative sequence homologues of these two gene products in NeleNPV (77% and 90% amino acid identity, respectively) and NeseNPV (44% and 63% amino acid identity, respectively) also encode a potential N-terminal signal and transmembrane domain (Garcia-Maruniak *et al.*, 2004; Lauzon *et al.*, 2004). Alternatively, the sawfly baculoviruses may not encode an envelope fusion protein or produce a budded virus since for almost all known non-lepidopteran NPVs virus replication occurs only in midgut epithelium (Federici, 1993).

NeabNPV repeats. Two repeat regions (*rep2* and *rep4*) each contain a cluster of localized repeat elements that are not interspersed throughout the genome (Figure 2.2A.). The remaining repeat clusters (*rep1*, *rep3*, *rep5*, *rep6* and *rep7*), however, have a common repetitive element with the consensus sequence 5'-CAACTTGTCAAATGTGTTGGACCTCGAGCCCAACAAACGCGACATCT -3' that is interspersed at multiple loci throughout the genome (Figure 2.2A and Figure 2.2B). This core sequence is present in NeleNPV direct repeat 5 but does not occur in the NeseNPV genome. The loci known as homologous regions, *hrs*, of the Lepidoptera-infecting NPVs have been found to function as enhancers of early gene transcription

(Guarino and Summers, 1986) and origins of DNA replication (Kool *et al.*, 1993; Pearson *et al.*, 1992). The *hrs* of the Lepidoptera-infecting NPVs have also been implicated as sites of DNA recombination (Hyink *et al.*, 2002). In NeabNPV the locations of this repeat element appears to correlate with major rearrangement events between the NeleNPV and NeabNPV genomes (Figure 2.3B).

Repeated Genes. Most lepidopteran baculovirus genomes, except *Spodoptera exigua* NPV (SeMNPV) (IJkel *et al.*, 1999), *Plutella xylostella* GV (PxGV) (Hashimoto *et al.*, 2000) and AdorGV (Wormleaton *et al.*, 2003), commonly encode members of a family of repeated genes known as *bro* genes. These genes are typically repeated elements of the baculovirus genomes and it has been proposed that *bro* gene products are involved in DNA replication and transcription (Zemskov *et al.*, 2000) as well as in maximizing polyhedra formation in specific host species (Bideshi *et al.*, 2003). Outside of the *Baculoviridae*, *bro* and *bro*-like genes have been found in ascoviruses, insect iridoviruses, entomopoxviruses (ALI gene family), the phycodnavirus ESV, bacteriophages and prophages integrated into bacterial genomes (Afonso *et al.*, 2001; Bawden *et al.*, 2000; Bideshi *et al.*, 2003; Jakob *et al.*, 200). Despite the widespread occurrence of *bro* genes, no homologues of these genes were found in NeabNPV or the other sawfly baculoviruses.

Five ORFs unique to NeabNPV and NeleNPV (*neab3*, *neab4*, *neab80*, *neab81* and *neab84*), however, share a common amino acid motif duplicated in two NeleNPV ORFs (Figure 2.4), and occur within the interspersed genome repeats (*rep1* and *rep7*). These repeated amino acid motifs are shared in two genes in NeleNPV but were not previously reported (Lauzon *et al.*, 2004). Duplicated genes flanked by repeated regions were

reported in NeseNPV (Garcia-Maruniak *et al.*, 2004), but they do not share the amino acid motifs of the NeabNPV and NeleNPV repeated genes.

The G+C composition of the putative NeabNPV ORFs within the repeat regions contrasts significantly with the genome average. While the predicted coding sequences of the NeabNPV genome had a mean G+C content of 34% ± 4%, eight putative ORFs had a G+C content greater than 40%, including four ORFs within the repeated segments, *neab3* (45% G+C), *neab4* (47% G+C), *neab80* (44% G+C) and *neab81* (46% G+C). In bacteria, the G+C content is homogenous throughout genomes and heterogeneous between genomes (Muto and Osawa, 1987). Furthermore, genome G+C content is related to phylogeny (Osawa *et al.*, 1992) and has been used to predict genes that have been acquired by horizontal gene transfer (Garcia-Vallve *et al.*, 2000a; Garcia-Vallve *et al.*, 2000b; Ochman *et al.*, 2005). As a result, G+C content is often used as a genome signature, and can be used to compare baculovirus genomes. Given that the base composition of the repeat region sequences contrast so greatly with the average of the sawfly baculovirus genomes and that the repeat regions are found in the NeabNPV and NeleNPV but not NeseNPV, it is possible that a common ancestor of NeabNPV and NeleNPV acquired those elements by horizontal transfer following its divergence from NeseNPV.

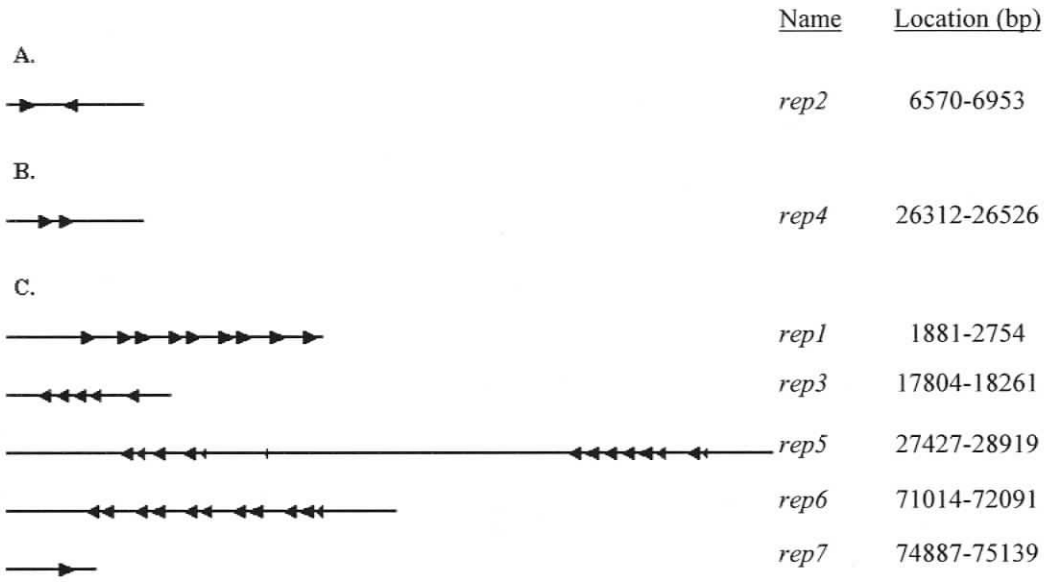


Figure 2.2. Loci of the NeabNPV genome with repeat elements. The triangles represent the repeated core element in the direction indicated by the point. The name and genome location of the repeat region is indicated on the right and the repeat element consensus sequences are (A) ATAAAAAACAGTAAATATT(C/T)CAATACGATGCAAACGCACGTGATTAATGT, (B) TCAG(A/C)ATTGTCGTTGTTGTTTT(G/C)AGTGTTTTCTGT(G/A)TTATTTC and (C) AGATGTCGCGTTTGTTGGGCTCGAGGTCCAACACATTTGACAAGTTG.

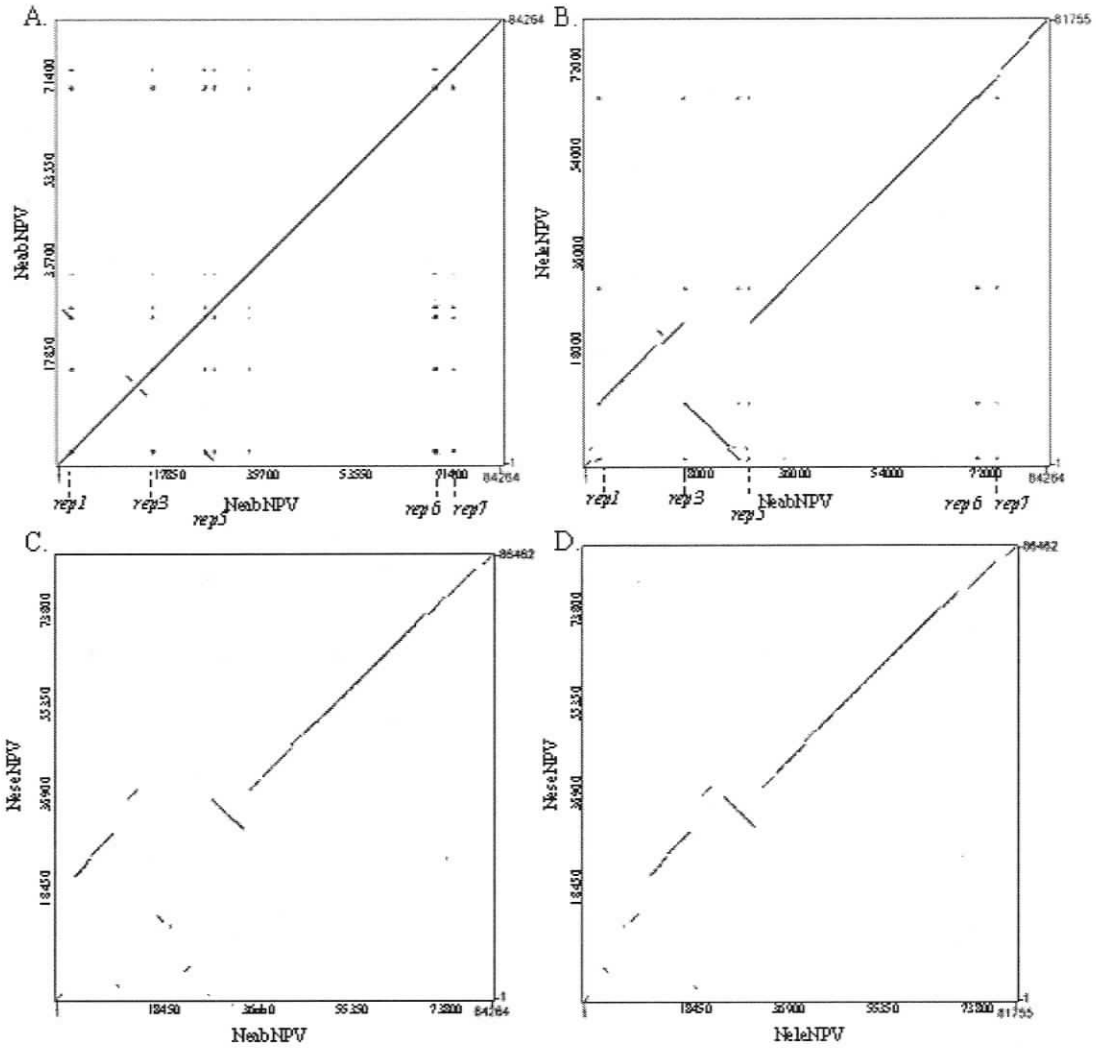


Figure 2.3. Genome parity plots. Genome plots were derived from comparing (A) NeabNPV against itself, (B) NeabNPV against NeleNPV, (C) NeabNPV against NeseNPV and (D) NeleNPV against NeseNPV. The axes are labeled with the genome name and the distance (in bp). The interspersed repeat sites of NeabNPV are labeled on plots A and B.

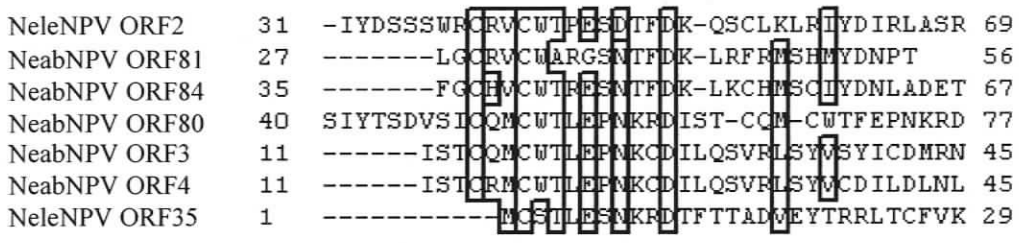


Figure 2.4. Alignment of conserved motif repeated in multiple predicted open reading frames (ORFs) of NeabNPV and NeleNPV genomes. Gaps in the alignment are indicated by a dash (-) and conserved amino acids are outlined.

NeabNPV Evolution. Sawfly baculoviruses represent a distinct taxon of the virus family *Baculoviridae* (Garcia-Maruniak *et al.*, 2004; Lauzon *et al.*, 2004). Nucleotide sequence plots revealed greater concordance between NeabNPV and NeleNPV than NeseNPV (Figure 2.3B and Figure 2.3C). Furthermore, NeabNPV and NeleNPV share the same conserved loci when compared to NeseNPV (Figure 2.3C and Figure 2.3D). Although all three host-species originally inhabited North America, *N. sertifer* is believed to have migrated via the Bering Land Bridge to Eurasia. Following this migration, *N. sertifer* became extinct in the New World only to be reintroduced c. 1925 (Lyons, 1964). This geographic isolation of host species is supported by the viral DNA sequences and is consistent with the divergence of NeseNPV from NeabNPV and NeleNPV.

The parity plots (Figure 2.3) illustrate that the region of the genome from 37-84Kb (*neab41-neab93*), relative to the *polyhedrin* gene, holds more concordance than the preceding region. Among the first 40 ORFs in this region of NeabNPV, only 67.5% have predicted sequence homologues in NeseNPV. Among ORFs 41-93 however, 87% have predicted sequence homologues. Similarly, among the first 40 ORFs only 6 (15%) ORFs are conserved in all sequenced baculovirus genomes, whereas among ORFs 41-93, twenty-three predicted ORFs (43.4%) are conserved in all sequenced baculovirus genomes.

Based on genome comparison, we propose a model for the evolution of sawfly baculoviruses whereby a repeat element appeared in the NeleNPV genome following its divergence from NeseNPV. This element underwent expansion and became interspersed throughout the genome. These interspersed repeats may have facilitated major genome rearrangements and the divergence of NeabNPV from NeleNPV.

Promoter Analysis. Gene expression of Lepidoptera-infecting baculoviruses is temporally regulated with four distinct stages: immediate-early, early, late and very late gene expression. Immediate-early and delayed-early genes are expressed before DNA replication and are transcribed by insect host RNA polymerase II (Fuchs *et al.*, 1983; Hoopes and Rohrmann, 1991). These early genes may encode promoter elements such as the TATA element that acts as an assembly site for host RNA polymerase II transcription complex (Guarino and Smith, 1992; Pullen and Friesen, 1995b) and the CAKT sequence element that enhances gene expression and serves as a transcription start site in early-expressed baculovirus genes (Blissard *et al.*, 1992; Pullen and Friesen, 1995b).

Putative TATA elements were identified upstream of 69 NeabNPV ORFs and CAKT elements were identified in the –40 to –20 bp region upstream of 12 ORFs. Based on the lepidopteran baculovirus model, ORFs with both elements should represent early-expressed genes. Six ORFs with putative TATA box motifs also have putative CAKT motifs. Among these are suspected homologues of the viral RNA polymerase subunit gene *lef-8* (Guarino *et al.*, 1998), *lef-11*, and *vlf-1* and four uncharacterized ORFs. CAKT was also identified in six predicted TATA-less promoters, including the early-expressed *dnapol* (Chaeychomsri *et al.*, 1995; Huang and Levin, 2001; Liu and Carstens, 1995; Tomalski *et al.*, 1988) and *p47* (Carstens *et al.*, 1993; Lapointe *et al.*, 2000) genes as well as the late-expressed *odv-e27* (Braunagel *et al.*, 1996b) and *vp39* genes (Blissard *et al.*, 1989; Heldens *et al.*, 1998; Thiem and Miller, 1989), and two uncharacterized ORFs. The occurrence of TATA-elements was significantly more frequent in the regions upstream of the predicted ORFs than in the whole genome, based on statistical

comparison with a random-locus data set, suggesting that it is a general transcription motif. The CAKT sequence was not significantly over-represented either 160 bp upstream, or in the –40 to –20 upstream regions, relative to the predicted translation start sites of NeabNPV ORFs. This may be due to the fact that the early genes would represent a small subset of viral genes and this element would not be a general transcription element. Alternatively, the absence of lepidopteran baculovirus immediate-early gene homologues in NeabNPV might suggest that sawfly baculoviruses utilize an alternate mechanism for early gene regulation.

Late gene expression in lepidopteran baculoviruses is associated with a DTAAG sequence motif (Morris and Miller, 1994). The DTAAG promoter element acts as the initiation site for transcription mediated by the viral-encoded RNA polymerase and the strength of expression from this site depends on the context of the sequence between it and the translational initiation site (Mans and Knebel-Morsdorf, 1998; Ooi *et al.*, 1989; Possee and Howard, 1987; Pullen and Friesen, 1995b; Weyer and Possee, 1988). The DTAAG motif was a common element in NeabNPV and was over-represented in the regions 160 bp upstream of the predicted ORF translation initiation sites.

Among the predicted NeabNPV ORFs that have well characterized putative sequence homologues in other baculoviruses, there is little evidence that the presence of these three promoter elements is indicative of the temporal expression of the genes. From a genome analysis perspective, the DTAAG sequence occurred in NeabNPV early genes as often as in late genes. However, it is likely that this site acts as an alternate start site for later transcripts. Whether the viral RNA polymerase of sawfly baculoviruses

utilizes a DTAAG motif sequence as a transcription start site, remains to be experimentally verified.

In order to explore alternate approaches to baculovirus promoter analysis we examined two features. The first approach involved determining whether there was a general over-representation of host transcription factor binding sites upstream of NeabNPV ORFs, and whether the occurrence of these binding sites were biased toward genes within a similar temporal group. We hypothesized that early baculovirus promoters would rely on host transcription factors to a greater degree than later genes, as they utilize the host transcription mechanism.

The second approach involved using AlignACE (Hughes *et al.*, 2000) to perform multiple alignments to identify conserved motifs. This approach successfully identified core promoter elements in *Drosophila* (Ohler *et al.*, 2002), and we hypothesized that it would identify over-represented conserved promoter motifs in the NeabNPV genome. In particular, we were interested whether the multiple alignment method would identify core promoter elements likely to be employed by the insect host cell transcriptional apparatus or common baculovirus elements.

To address the first question regarding the utilization of known host transcription factor binding sites, the upstream sequences were queried against the TRANSFAC database. Three experimentally confirmed insect promoter elements were over-represented in the upstream sequences of predicted NeabNPV ORFs. One was the TATA-box, a common core promoter element common in insect and baculovirus genomes. Another was the motif TGAGHB that corresponds to the DNA binding motif for the zeste protein.

Zeste is a transcription factor that facilitates transvection, the mechanism by which a gene can *trans*-regulate the expression of a homologue (Duncan, 2002; Kennison and Southworth, 2002; Wu and Morris, 1999). It has been observed that AcMNPV *hr1* can act as an enhancer in *trans* in mammalian cells (Viswanathan *et al.*, 2003). Although the *hr1* region did not permit transvection of genes encoded by host insect genomes, it is possible that hymenopteran baculovirus genes could utilize zeste-mediated *trans*-regulation. The zeste promoter motif has been identified upstream of seven genes with a bias toward early-expressed genes, including *dnapol* (Chaeychomsri *et al.*, 1995; Liu and Carstens, 1995; Huang and Levin, 2001; Tomalski *et al.*, 1988), *iap-3* (Carpes *et al.*, 2005; Ikeda *et al.*, 2004), *p40* (Lu *et al.*, 1996), *p47* (Carstens *et al.*, 1993; Lapointe *et al.*, 2000) and *38K* (*pp31*) (Smith *et al.*, 1982). Two genes with this sequence upstream, *pif* and *vp91*, are expressed late in lepidopteran NPVs (Gutierrez *et al.*, 2004; Russell and Rohrmann, 1997; Simon *et al.*, 2005). Zeste promoter motifs were identified in the upstream-sequences of a further 13 predicted NeabNPV ORFs.

A CREB/ATF47 binding site (known as the CRE element) mediates transcription in response to members of the CREB/ATF protein family. This is a large family of leucine zipper transcription factors (Hai *et al*, 1989) that, despite their diverse activities, share a common ability to respond to environmental signals and maintain cellular homeostasis. Some ATFs (ATF2, ATF3 and ATF6) play a role in mediating stress response (Hai *et al*, 1999); CREB and ATF1 regulate transcription in response to intracellular cAMP concentrations (Hai and Hartman, 2001); and ATF4 acts as a negative regulator of CRE-dependent transcription (De Cesare and Sassone-Corsi, 2000). ATF proteins are important elements in regulating early transcription of the *Adenoviridae*. The adenovirus

E1A protein interacts with host-encoded ATF-2 to activate transcription of several adenovirus early genes (Liu and Green, 1990). In the NeabNPV genome, the CREB/ATF binding sites were observed upstream of both the putative early-expressed gene *dnapol* (Chaeychomsri *et al.*, 1995; Huang and Levin, 2001; Liu and Carstens, 1995; Tomalski *et al.*, 1988) and *lef-1* (Passarelli and Miller, 1993) as well as the putatively late-expressed gene *pif* (Gutierrez *et al.*, 2004; Simon *et al.*, 2005). The remaining six ORFs with upstream CREB/ATF binding sites did not have functionally characterized orthologues in other baculoviruses.

Using the second approach, alignment of upstream sequences identified nine sequences that we denoted "M1" to "M9" (Table 2.1 and Table 2.2). Contrary to our prediction, promoter alignment did not report the core *Drosophila* promoter elements or the known baculovirus promoter elements. Within these over-represented conserved motifs, however, certain spatial and temporal patterns were observed.

M1 motif was not over-represented in genes known to be temporally distinct, but rather was clustered spatially. Of the 10 occurrences of this motif, eight (8) of them are within a 30 Kb region (747-29,973 bp), within the more variable half of the circular NeabNPV genome. Similarly, 5 of 6 ORFs with the upstream M2 motif were within an approximately 5 Kb region (82,641-84,264 bp and 0-3,558 bp). The M8 motif was upstream of the late-expressed *polh* (Choi *et al.*, 1998; Chou *et al.*, 1996; Rohel and Faulkner, 1984; van Strien *et al.*, 1992) and *gp41* (Liu and Maruniak, 1995; Whitford and Faulkner, 1992) genes as well as 5 other putative ORFs. Five of the seven ORFs with an upstream M8 motif also possess putative TATA and DTAAG motifs.

Table 2.2. Predicted putative promoter elements for NeabNPV.

A. EPD Analysis[a]

Factor	Consensus Sequence	p-value[d]
TATA Box	TATA	0.114(0.041)[e]
	TATAA	0.000(0.016)[e]
	TATAT	0.041(0.305)[e]
Zeste	TGAGHB	0.044
CREB/ATF-47	ACGTCA	0.028

B. Over-represented Elements[b]

Motif	Pictogram	p-value[d]
M1	 TGTCRVDWDYRTWNRR	0.002
M2	 RACCYTSRM	0.047
M3	 WNNNNWTTTNNHAHCA	0.001
M4	 CAMATAWWT	0.002
M5	 TTRWNTAYWT	0.047
M6	 WCDWTNNRCAA	0
M7	 WNNTNHADAMRAY	0.005
M8	 TTWTNWTTNNTA	0.026
M9	 HACANDWTTG	0.001

C. Common Baculovirus Elements[c]

Motif	Consensus Sequence	p-value[d]
TATA Box	TATA	0.114(0.041)[e]
CAKT	CAKT	0.36 (0.623)[e]
DTAAG	DTAAG	0

[a] Transcription factors derived from SignalScan analysis of the Eukaryote promoter database (see materials and methods). Only those that occur at a significantly more often in the upstream sequence than the whole genome were reported.

[b] AlignAce analysis of upstream sequences identified a number of putative common elements. Only those that occur at a significantly more often in the upstream sequence than the whole genome were reported.

[c] Common baculovirus promoter elements.

[d] The p-value represents a statistical test that the element occurs more often –160 bp upstream of predicted ORFs than in the whole NeabNPV genome.

[e] Value in brackets represents a statistical test that the element occurs more often –40 to –20 bp region upstream of predicted ORFs than in the whole NeabNPV genome.

The motifs M2 and M8 occurred in only 6 and 7 upstream sequences, respectively, but these are still significantly more often than in the whole genome. The motif M3 occurred upstream of 10 predicted ORFs including: *p33*, *trypsin-like*, *lef-9*, *p74*, *pif-2*, and *rcc1*. There is no available temporal characterization for all of these genes except *lef-9* and *p74*, which are both late-expressed (Acharya and Gopinathan, 2002; Guarino *et al.*, 1998; Kuzio *et al.*, 1989; Rashidan *et al.*, 2005). The other motifs (M4, M5, M6, and M7) did not appear to have characteristic clustering or temporal properties.

Our hypothesis that host transcription factor binding sites might be enriched in the promoters of early-expressed baculovirus genes appears to hold for the zeste protein binding site but not for the CREB/ATF47 binding site. Conversely, motifs characterized by promoter alignment with AlignACE did not correspond with temporally grouped genes, however the spatial clustering of genes with these motifs is intriguing. These patterns and hypotheses require experimental confirmation and emphasize the need for more sophisticated analyses of baculovirus promoters.

Concluding remarks. Although this study represents the third sawfly baculovirus genome to be completely sequenced, our analysis conforms with previous observations that the sawfly baculoviruses represent a distinct taxon (Garcia-Maruniak *et al.*, 2004; Lauzon *et al.*, 2004; Herniou *et al.*, 2004). Our study suggests a possible mechanism for sawfly baculovirus evolution through genome rearrangement events between interspersed genome repeats. Rearrangement between these repeats may be similar to the genome arrangements described between the repeated lepidopteran *hr* regions. Although not as well conserved as the lepidopteran

baculovirus *bro* genes, the predicted genes encoded within the NeabNPV repeats share a common amino acid motif.

Our comparison of gene content and genome rearrangement agrees with the hypothesis proposed by Herniou *et al.* (2004), that baculoviruses coevolve with their host. We observed that the baculovirus of geographically isolated Old World *N. sertifer* (NeseNPV) displayed greater divergence from the baculoviruses of New World sawflies, *N. abietis* (NeabNPV) and *N. lecontei* (NeleNPV).

The intriguing absence of immediate early gene homologues and membrane fusion proteins in NeabNPV is consistent with previous studies of non-lepidopteran baculoviruses (Afonso *et al.*, 2001; Garcia-Maruniak *et al.*, 2004; Lauzon *et al.*, 2004), and raises interesting questions regarding how early gene expression is mediated and how the sawfly baculovirus infection process differs from that of the lepidopteran baculoviruses. These questions, along with the renewed interest in sawfly baculoviruses in pest management, provide us with an opportunity to examine the unique and exciting characteristics of the baculoviruses outside the order *Lepidoptera*.

Chapter 3. *In vivo* replication kinetics and transcription patterns of the balsam fir sawfly, *Neodiprion abietis*, nucleopolyhedrovirus.

3.1. Abstract

Baculoviruses utilize a temporal cascade of gene expression to coordinate efficient viral propagation. Viral DNA replication and transcription of lepidopteran baculoviruses has been well characterized due to the availability of *in vitro* viral propagation systems. In contrast, almost nothing is known about viral replication and gene expression of non-lepidopteran NPVs. *In vitro* propagation systems are not available for dipteran or sawfly baculoviruses. To address this, *in vivo* DNA replication kinetics and transcription of the baculovirus pathogenic for the hymenopteran *Neodiprion abietis* (NeabNPV) were investigated. Transcription of selected NeabNPV genes appeared to follow a cascade of early, late and very late expression, similar to lepidopteran baculoviruses. Transcription of NeabNPV *lef-1*, *lef-2* and *dna polymerase* genes was observed as early as 2 hours post infections (hpi). Late gene expression was marked by the transient transcription of putative viral RNA polymerase subunits, *lef-8* and *lef-9*, at 6 hpi. All late genes were transcribed from 6 hpi. The *vlf-1* gene, which encodes the transactivator responsible for very late gene expression, was observed from 24 hpi, but the very late *polh* transcript was not observed within 72 hpi. This suggests that *vlf-1* gene expression alone is not sufficient for *polh* transcription in NeabNPV. In contrast to the lepidopteran baculoviruses, transcription of the *pif* gene was observed from 2 hpi, suggesting that it is expressed early in NeabNPV. Viral DNA replication kinetics support this time-line

for NeabNPV infection with an increase in viral DNA copies over the 72 hpi to a predicted maximum at 79 hpi. This study provides the first insight into the DNA replication and gene expression of a non-lepidopteran baculovirus.

3.2. Introduction

The Baculoviruses are a family of invertebrate viruses best characterized in insects of the order *Lepidoptera, Diptera*, and *Hymenoptera*. Baculoviruses have relatively large DNA genomes encoding between 80 and 180 genes. Molecular characterization has shown a strong bias toward the lepidopteran baculoviruses, which coordinate gene transcription via a temporally regulated cascade of transcription events: immediate early, early, late and very late.

The majority of studies characterizing baculovirus viral DNA replication and gene transcription have been conducted *in vitro*, in cultured insect cells. Currently, there are no *in vitro* propagation systems for non-lepidopteran baculoviruses, so their infection processes remain largely unknown. Recent genome sequence analyses (Afonso *et al.*, 2001; Garcia-Maruniak *et al.*, 2004; Lauzon *et al.*, 2004; This thesis, Chapter 2) have revealed intriguing differences in the early-expressed genes of lepidopteran versus non-lepidopteran baculoviruses. In lepidopteran baculoviruses, the first genes transcribed are the immediate early genes, whose gene products transactivate other virus-encoded genes. The immediate early genes, ubiquitous and conserved within the lepidopteran baculoviruses, are curiously absent in the non-lepidopteran baculoviruses, based on amino acid sequence homologies of encoded gene products.

The immediate early and early genes of lepidopteran baculoviruses utilize the host cell RNA polymerase II and transcription factors for their expression. The immmediate early gene products typically act as strong transactivators to specifically upregulate transcription of the early and late baculovirus genes. The early genes

encode gene products involved in DNA replication and/or transcription of virus-encoded genes. A viral α-amanitin resistant RNA polymerase is employed to transcribe late genes (Fuchs *et al.*, 1983; Grula *et al.*, 1981; Guarino *et al.*, 1998; Huh and Weaver, 1990), many of which are involved in the formation of the budded virion and occluded virion structures. The late-expressed *vlf-1* (very late expression factor-1) gene encodes a transactivator that upregulates transcription of the very late genes, such as those encoding the polyhedrin and p10 proteins (McLachlin and Miller, 1994; Yang and Miller, 1999). Very late gene products are primarily involved in virus occlusion and cell lysis.

Table 3.1 summarizes the temporal class, function, and earliest observed time of expression *in vitro* of a variety of lepidopteron baculovirus genes. The observed temporal expression pattern among baculoviruses is fairly consistent. Among lepidopteran baculoviruses, immediate early gene expression appears to occur within 3 hours post infection (hpi), followed by early genes involved in viral DNA replication and transcription ranging between 2-12 hpi. The subunits of viral RNA polymerase are synthesized from 6-12 hpi, permitting the expression of late genes from 6-24 hpi. The *vlf-1* gene product transactivates the *polh* (polyhedrin) and *p10* genes by 72 hpi. This transactivation is considered the defining transition between late and very late gene expression.

Little is known about the coordination of gene transcription in non-lepidopteran baculoviruses, because *in vitro* viral propagation systems are not currently available. The present study addresses this by investigating viral DNA replication and gene expression of the balsam fir sawfly, *Neodiprion abietis* (*Hymenoptera*)

nucleopolyhedrovirus (NeabNPV). We use reverse transcriptase PCR (RT-PCR) to map virus temporal gene expression, and quantitative PCR (QPCR) to examine viral DNA replication kinetics, in NeabNPV infected *N. abietis* larvae.

Table 3.1. Class, function, and earliest observed time of expression *in vitro* for lepidopteran baculovirus genes.

Class and Function	Gene Name	Earliest Transcribed (hpi)[a]	References
Immediate Early – Transactivation			
	ie-0	0.25	(Kovacs *et al.*, 1991)
	ie-1	0.25-3	(Carstens *et al.*, 2002; Kovacs *et al.*, 1991; van Strien *et al.*, 2000)
	ie-2	0.5-3	(Carson *et al.*, 1988; Carstens *et al.*, 2002; Theilmann and Stewart, 1992)
	pe38	1-3	(Carstens *et al.*, 2002; Krappa and Knebel-Morsdorf, 1991; Wu *et al.*, 1993)
	me-53	3	(Wang *et al.*, 2004)
Early – DNA Replication			
	dbp	6	(Mainz *et al.*, 2002)
	lef-1	3	(Passarelli and Miller 1993)
	lef-2	2, 12-18	(Ahrens and Rohrmann, 1995; Mainprize *et al.*, 1986; Sriram and Gopinathan, 1998)
	lef-3	3-6	(Ahrens and Rohrmann, 1995; Chen *et al.*, 2004; Li *et al.*, 1993)
	dnapol	2-6	(Chaeychomsri *et al.*, 1995; Passarelli and Miller, 1993; Liu and Carstens, 1995; Huang and Levin, 2001; Tomalski *et al.*, 1988)
	helicase	4-5	(Bideshi *et al.*, 1998; IJkel *et al.*, 1999)
Early - Transcription			
	pp31	4	(Smith *et al.*, 1982)
	lef-5	12	(Heldens *et al.*, 1998)
	lef-6	4	(Lin and Blissard, 2002)
	lef-11	4	(Lin *et al.*, 2001)
	lef-12	6	(Lapointe *et al.*, 2000)

[a]hpi = hours post infection.

Table 3.1. cont.

Class and Function	Gene Name	Earliest Transcribed (hpi) [a]	References
RNA Polymerase Subunits			
	lef-4	3-8	(Durantel *et al.*, 1998; Heldens *et al.*, 1998)
	lef-8	12	(Acharya and Gopinathan, 2002)
	lef-9	6-12	(Acharya and Gopinathan, 2002; Guarino *et al.*, 1998)
	p47	6	(Carstens *et al.*, 1993; Lapointe *et al.*, 2000)
Late - Structural			
	vp1054	6	(Olszewski and Miller 1997)
	gp41	12	(Liu and Maruniak, 1995; Whitford and Faulkner, 1992)
	vp39	8-24	(Blissard *et al.*, 1989 Heldens *et al.*, 1998;; Thiem and Miller, 1989)
	p6.9	16	(Wang *et al.*, 2001)
	p74	24	(Kuzio *et al.*, 1989; Rashidan *et al.*, 2005)
	odv-e56	16-24	(Braunagel *et al.*, 1996a; Rashidan *et al.*, 2005)
Very Late - Transcription			
	vlf-1	12	(McLachlin and Miller, 1994)
Very Late - Structural			
	polh	12-96	(Choi *et al.*, 1998; Chou *et al.*, 1996; McLachlin and Miller, 1994; Rohel and Faulkner, 1984; van Strien *et al.*, 1992)
	p10	72	(Faktor *et al.*, 1997; Razuck *et al.*, 2002)

[a] hpi = hours post infection.

3.3. Materials and Methods

Virus Source. NeabNPV was isolated in 1997 from *N. abietis* larvae collected between Stephenville (latitude 48°32' N, longitude 58°33' W) and Corner Brook, Newfoundland, Canada. The virus was purified as described by Moreau *et al.* (2005) and as described in the 'Methods and Materials' section of chapter 2.

Larval infection. Balsam Fir sawfly larvae (*N. abietis*) were collected from balsam fir (*Abies balsamea*) branches at the Old Man's Pond region (Corner Brook) of Western Newfoundland, Canada, in July of 2003. The larvae were maintained at 4°C on balsam fir branches until they were harvested. The larvae were starved for 12-15 hours at room temperature prior to being inoculated. Head capsules of the larvae were measured, according to the procedure described by Carroll (1962) and only second and third instar larvae were selected for experimentation. NeabNPV polyhedral inclusion bodies (PIBs) were mixed with a 10% pasteurized liquid honey solution and *N. abietis* test larvae were each inoculated by feeding on a 5 µl drop of honey containing an estimated 10,000 PIBs. Control larvae were inoculated by feeding on 5 µl of a 10% pasteurized liquid honey solution with no PIBs. Larvae were maintained on balsam fir foliage, which had been decontaminated with 5% bleach for 30 minutes and rinsed with water.

Larvae used in this experiment were sacrificed at 0.5, 1, 2, 6, 12, 24, and 72 hours following inoculation. Each sacrificed larva was washed for 60 seconds in 5% bleach and rinsed for 60 seconds in DEPC-treated water. The head and tail were removed,

and the larval gut was separated from the fat body and cuticle in sterile phosphate buffered saline (pH 7.4) solution. The food bolis was removed by removing the peritrophic membrane and the dissected gut was stored in 1 ml of RNALater (Ambion Inc) at -20 °C.

Viral RNA and DNA isolation. Each excised larval midgut was homogenized in 50 µl TRIzol reagent (Invitrogen) containing sand (ignited and baked 4 hrs). The ground midgut suspension was collected and TRIzol reagent (Invitrogen) was added to 1 mL. Debris was collected by centrifugation at 12,000 × g for 10 minutes at 4°C. Each homogenized sample was incubated for 10 minutes at 25°C. Each sample was extracted with a 1/5 volume of chloroform. Tubes were mixed for 15 seconds and incubated 2 minutes at 25°C. The TRIzol and phenol-chloroform phases were separated by centrifugation at 12,000 × g for 15 minutes at 4°C and the aqueous phase was collected. RNA was precipitated from the solution by addition of ½ volume of isopropyl alcohol, incubated for 10 minutes at 25°C, and centrifuged at 12,000 × g for 10 minutes at 4°C. The supernatant was removed and the pellet was washed once with 1 mL 75% ethanol. The sample was mixed by vortex and centrifuged at 7,500 × g for 5 minutes at 4°C. The pellet was dried and resuspended in RNAse-free water (Invitrogen) by incubation at 60°C for 10 minutes. RNA samples were stored at -80°C.

DNA was precipitated from the interphase and organic phase with 1/3 volume 100% ethanol and incubated for 3 minutes at 25°C. Precipitate was collected by centrifugation at 2,000 × g for 5 minutes at 4°C and removal of supernatant. The

pellet was dried and resuspended in RNAse-free water (Invitrogen) by incubation at 60°C for 10 minutes. DNA samples were stored at -80°C.

RNA Reverse Transcription PCR. The RNA samples were treated with 3 U of DNAse (Invitrogen) in a buffered solution (20 mM Tris-Cl, pH 8.4; 50 mM KCl; 2 mM $MgCl_2$) for 15 minutes at 25°C. The reaction was terminated by addition of EDTA to 2.5 mM and incubation at 65°C for 10 minutes. First strand cDNA synthesis was performed by incubating 50 ng DNAse-treated total RNA with 200 U Superscript II Reverse Transcriptase in 20 µl PCR buffer (50 mM KCl; 25 mM Tris-HCl, pH 8.3; 500 µM dNTP; 10 µM DTT; 10 U RNAse Inhibitor) at 25°C for 10 minutes. The reaction was then incubated at 42°C for 50 minutes and inactivated by incubation at 70°C for 15 minutes. The synthesized cDNA was stored at -20°C.

The primers used to amplify the temporal standards are listed in Table 3.2. The amplicons were designed to yield products of 200-300 bp in length. A positive control (NeabNPV genomic DNA), a negative control (cDNA from uninfected larva), and a host insect control (28S rRNA gene-specific primers) were included in each PCR set. The DNA templates were amplified using gene specific primers and Platinum *Taq* polymerase (Invitrogen), using the manufacturer's protocol, with temperature cycling between 95°C (9 mins.), followed by 45 cycles of 95°C (30 secs.), 55°C (60 secs.), and 72°C (90 secs). All products were resolved by gel electrophoresis in 2% agarose and stained with SYBR Green.

Table 3.2. Primers used in reverse transcription and real time PCR of RNA extracted from *N. abietis* larval midguts infected with NeabNPV.

Primer Target	Expected Size (bp)	Forward Primer Sequence (5' to 3')	Reverse Primer Sequence (5' to 3')
actin	900	AGGAGACGAGGGCACAGAGCAAAAG	TTGGAAGGTAGAGAGGGAGGCGAG
28S rDNA	153	AAAGATCGAATGGGGAGATTCATC	CGTCCTACTAGGGGAGAAGTGCAC
dna pol	218	CAGCGCACAGTGGACGTGGTT	GCCGTTTGGGAAATTGTTGCA
lef1	223	GCGATGCAACCGACGACGATA	TTGTATTGCGCTTCGTCAATTTTG
lef2	259	ATGGATAAATGCAAATACGTCAAAA	CGGCGGTCCTTGTTGTTGATC
lef-8	254	GCCGGCATTTGTATGTTGGGA	CCGGTAAAATGATCGAATTTGCA
lef-9	195	ACCGAACATCGGCGAACTGAT	CGGCGATCGAAAATTGTGCAA
gp41	206	TCCATTCCGTTGAGCGCAAAA	TCGCGCTCTTGTTCGATTATCC
p74	320	GAATTCGCACCACCGAAGCCACCGT	CTGCAGGGCGGTTGCCCATTCA
vlf1	216	GGGGTGTACGAATTCGCCATC	CAAAATTTCTTTCCGTCAAAACGTT
polh	176	GCAGGTTATCAAACCTCGGCC	GGACCCGTTCGTGGGTCCAAC
pif	188	CGGACATTTTGAAGCTACCGTCG	CGGCGGAATTACAAGAGTCGTCG

Quantitiative PCR. Viral DNA was quantified by real time PCR amplification (QPCR) with primers specific to the *polh* gene (Table 3.2). The viral copy number was then normalized against host cell genome copy number by QPCR with primers specific to the *28S* and *actin* genes (Table 3.2). The *polh* gene was amplified by PCR and cloned into pGem-T (Promega). Plasmid DNA concentration was quantified using a ND-1000 spectrophotometer (NanoDrop) and dilution standards were generated ranging from 2.5×10^5 to 4×10^2 copies. For each standard dilution, QPCR was performed using *polh*-specific primers and a standard curve was generated. For each larval DNA extract, QPCR was performed using *polh*- and *actin*-specific primers. The viral DNA was quantified and *polh* amplicon was normalized against host *28S* and *actin* to derive the number of viral copies per cell. The reported number of viral copies represented the mean of two larvae at each time point. The DNA templates were amplified with an Mx4000 thermocycler (Stratagene) using gene specific primers (0.67μM) and 20 μM dNTP in MX4000 QPCR Buffer (Stratagene), with Platinum *Taq* polymerase (Invitrogen). Temperature cycling between 95°C (9 mins.) followed by 40 cycles of 95°C (15 secs.), 55°C (30 secs.), and 72°C (45 secs) was used. All products were confirmed by resolving by gel electrophoresis in 2% agarose and stained with SYBR Green.

3.4. Results

Transcription of predicted early baculovirus genes. The earliest-expressed baculovirus genes with sequence homologues in the NeabNPV genome include the DNA polymerase, encoded by *dnapol*, and the DNA primase, encoded by *lef-1* (Chaeychomsri *et al.*, 1995; Passarelli and Miller, 1993; Liu and Carstens, 1995; Huang and Levin, 2001; Tomalski *et al.*, 1988). The *lef-2* gene has also been observed as an early expressed gene (Mainprize *et al.*, 1986) and its product has been shown to interact with the early expressed *lef-1* gene product, DNA primase (Evans *et al.*, 1997). However, l*ef-2* is known also as a late-expressed gene that stimulates transcription from late promoters (Passarelli and Miller, 1993; Sriram and Gopinathan, 1998). The NeabNPV *lef-1*, *lef-2*, and *dnapol* transcripts were all first observed as early as 2 hpi (Figure 3.1). The *lef-2* transcript was not observed at 6 hpi but was detected again from 12 to 72 hpi, suggesting that its transcription might be suppressed between early and late transcription.

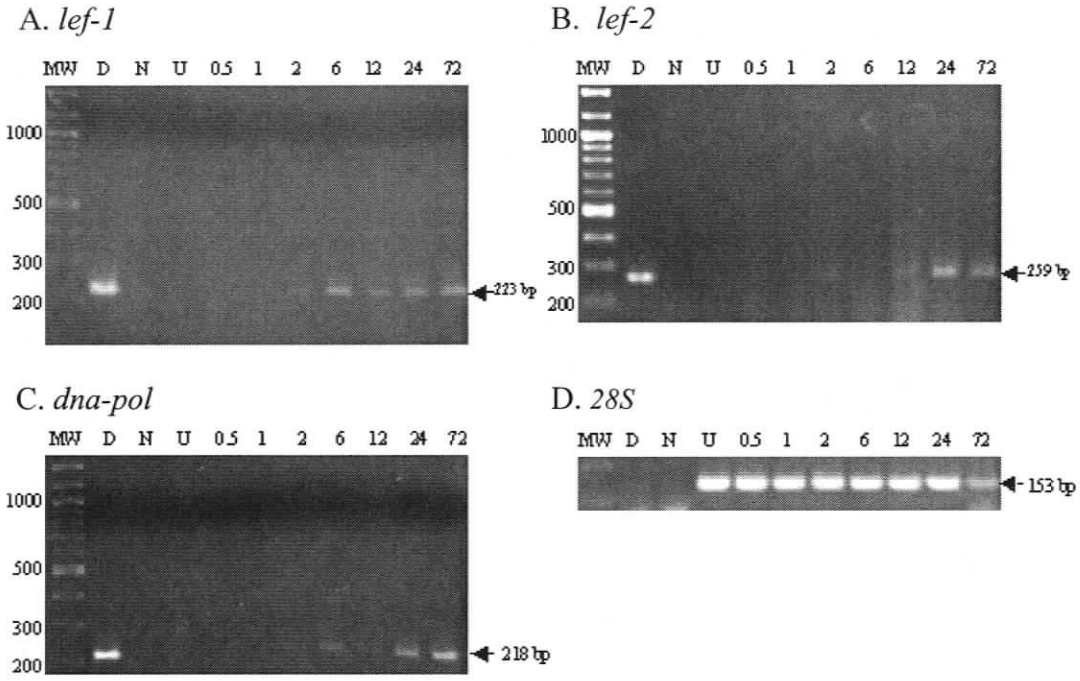


Figure 3.1. RT-PCR analysis of putative early genes compared with a host gene. (A) *lef-1*, (B) *lef-2*, (C) *dnapol*, and (D) host cell *28S* rRNA gene. RT-PCR was performed from RNA extracted from uninfected and NeabNPV-infected larvae. The expected size of each RT-PCR product is indicated to the right of each panel. Lane D) amplification from NeabNPV DNA template (positive control); Lane N) amplification from uninfected larva (negative control). The remaining lanes are amplicons from RNA extracted from larvae 0.5 to 72 hpi, as indicated.

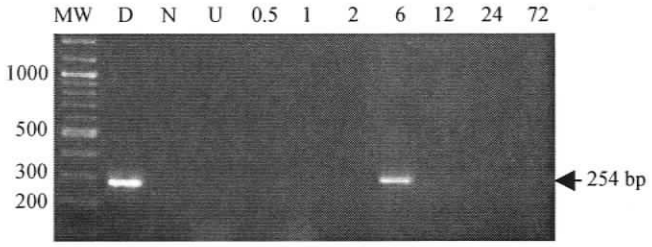
Transcription of genes encoding RNA polymerase baculovirus subunits. The transcripts for *lef-8* and *lef-9*, together with *lef-4* and *p47*, comprise the four subunits of the virus-encoded RNA polymerase (Guarino *et al*., 1998). Previous studies of *Autographa californica* NPV (AcMNPV) (Guarino *et al*., 1998) and *Bombyx mori* NPV (BmMNPV) (Acharya and Gopinathan, 2002) *lef-8* and *lef-9* gene expression identified transcripts as early as 6 and 12 hpi. In NeabNPV, the transcripts for *lef-8* and *lef-9* were only identified at 6 hpi (Figure 3.2), indicating that this transcript may be expressed transiently. Our data is consistent with *lef-8* and *lef-9* expression at 6-12 hpi in AcMNPV and BmMNPV, except that we did not detect NeabNPV *lef-8* and *lef-9* transcription at 12 hpi or later time points.

Transcription of predicted late and very late baculovirus genes. Transcription two putative late-expressed genes, *gp41* and *p74* (associated with viral ODV) and two putative very late expressed genes, *vlf-1* and *polyhedrin* (*polh*), transcripts were examined (Figure 3.3). While the *gp41* and *p74* transcripts were observed after 6 hpi the very late factor gene, *vlf-1*, transcript was observed at 24 and 72 hpi (Figure 3.3C). Transcripts for the *polh* gene were not observed within 72 hpi (Figure 3.3D). Transcriptional analysis in AcMNPV observed that *polh* and *vlf-1* transcription were co-ordinately transcribed from 12-48 hpi (McLachlin *et* al., 1994). NeabNPV *vlf-1* transcription was not observed until 24 hpi and *polh* transcripts were not identified at all in the 72 hpi. This suggests that the very late gene expression of NeabNPV may be delayed relative to lepidopteran baculoviruses. The *pif* gene transcript was first observed at 2 hpi and until 72 hpi (Figure 3.4). This contrasts with observations in

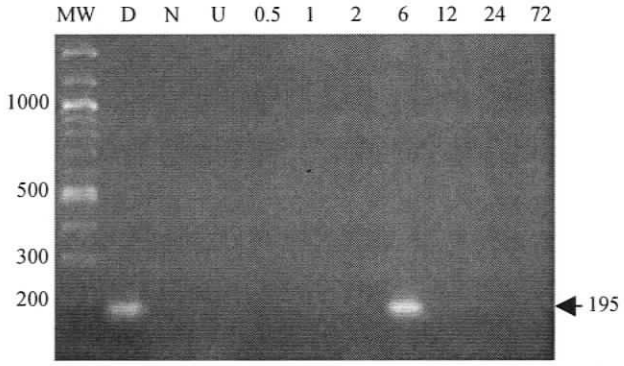
lepidopteran baculoviruses which demonstrate that the *pif* gene is only expressed late in infection (Gutierrez *et al.*, 2004; Simon *et al.*, 2005).

NeabNPV replication kinetics. Previous studies (Knudson and Tinsley, 1978) have indicated that 88% of lepidopteran baculovirus DNA synthesis occurs 8-12 hpi in *Spodoptera frugiperda* cell lines (SF-21). Using quantitative PCR (QPCR), a similar trend was observed in the SF-9 cell line (Rosinski *et al.*, 2002) with an exponential increase in viral DNA synthesis up to 20 hpi. Standard curves were generated for QPCR quantification of *N. abietis* DNA using *actin*-specific primers and NeabNPV DNA using *polh*-specific primers (Figure 3.5). A linear correlation was achieved for both standard curves. Quantification of NeabNPV DNA following *in vivo* infection showed a different trend (Figure 3.6) with a consistent polynomial increase in viral DNA copy number to approximately a predicted 2.1×10^5 vDNA copies per cell at 79 hpi (t_{max}).

A. *lef-8*



B. *lef-9*



C. *28S*

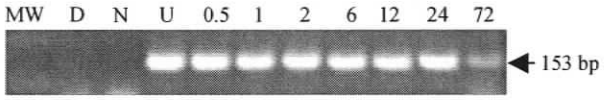


Figure 3.2. RT-PCR analysis of putative RNA polymerase subunit genes compared with a host gene. (A) *lef-8*, (B) *lef-9*, and (C) host cell *28S* rRNA gene. RT-PCR was performed using RNA extracted from uninfected and NeabNPV-infected larvae. The expected size of each RT-PCR product is indicated to the right of each panel. Lane D) represents amplification from NeabNPV DNA template (positive control); Lane N) uninfected larva (negative control). The remaining lanes are amplicons from RNA extracted from larvae 0.5 to 72 hpi, as indicated.

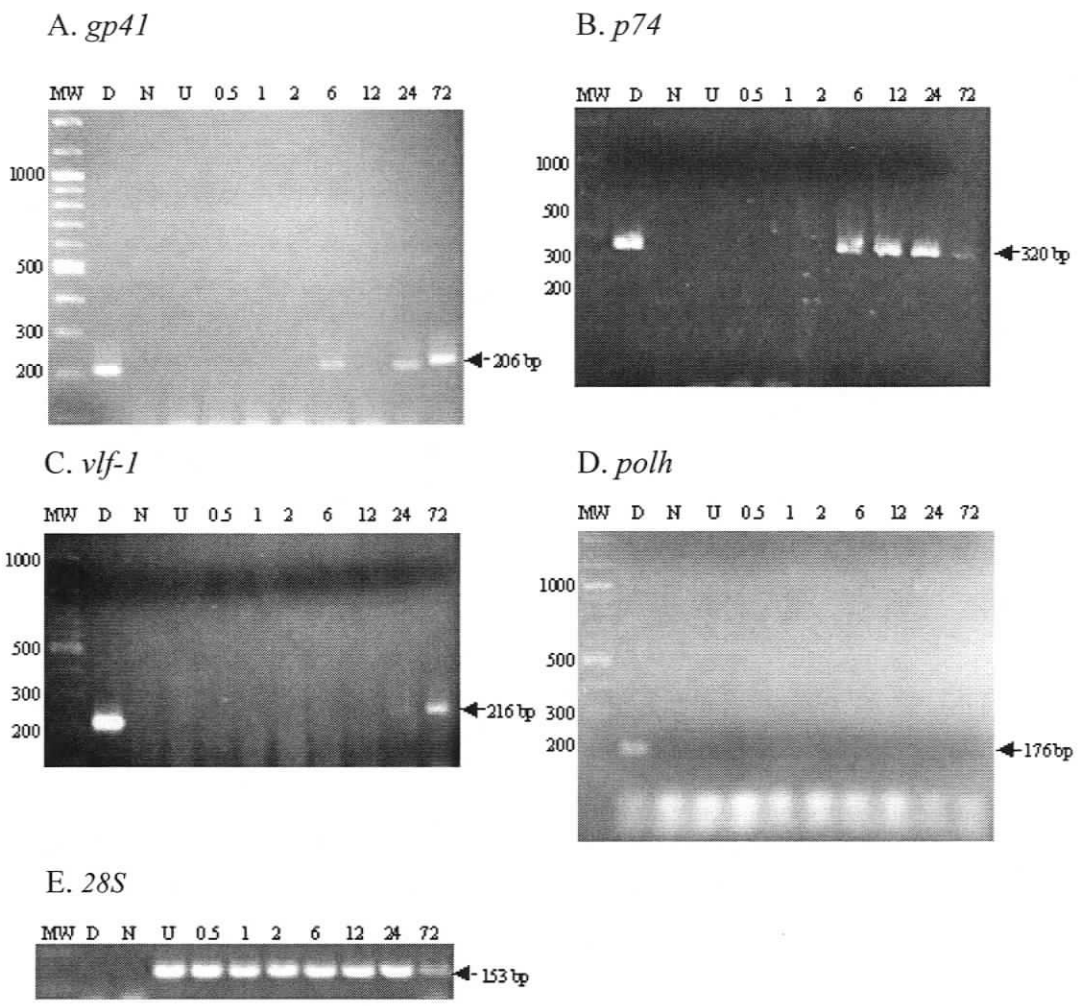


Figure 3.3. RT-PCR analysis of putative late and very late genes compared with a host gene. (A) *gp41*, (B) *p74*, (C) *vlf-1*, (D) *polh*, and (E) host cell *28S* rRNA gene. RT-PCR was performed from RNA extracted from uninfected and NeabNPV-infected larvae. The expected size of each RT-PCR product is indicated to the right of each panel. Lane D) amplification from NeabNPV DNA template (positive control); Lane N) amplification from uninfected larva (negative control). The remaining lanes are amplicons from RNA extracted from larvae 0.5 to 72 hpi, as indicated.

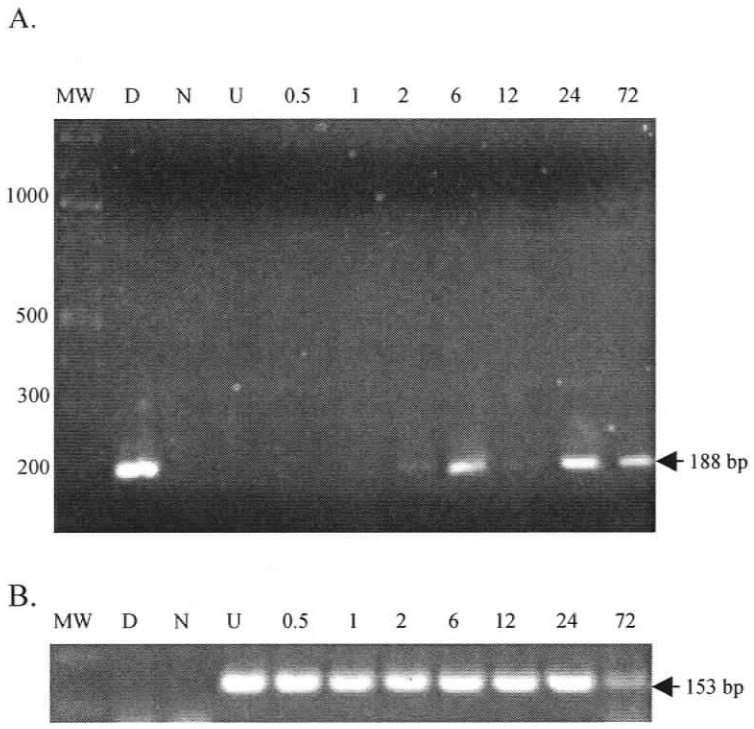
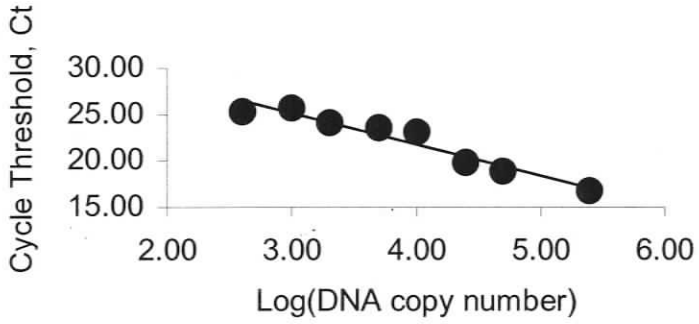


Figure 3.4. RT-PCR analysis of putative *pif* gene compared with a host gene. (A) *pif-1* and (B) host cell *28S* rRNA gene. RT-PCR was performed from RNA extracted from uninfected and NeabNPV-infected larvae. The expected size of each RT-PCR product is indicated to the right of each panel. Lane D) amplification from NeabNPV DNA template (positive control); Lane N) amplification from uninfected larva (negative control). The remaining lanes are amplicons from RNA extracted from larvae 0.5 to 72 hpi, as indicated.

A.



B.

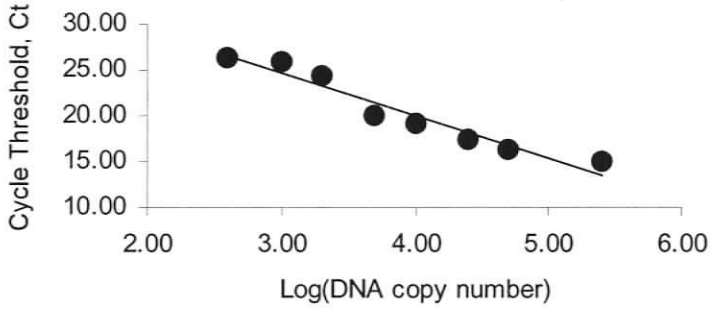


Figure 3.5. Calibration curves for quantitative PCR of *N. abietis* insect genomic DNA and NeabNPV DNA. (A) *N. abietis*-specific actin and (B) NeabNPV-specific *polh* genes were cloned and serial dilutions of the clones were amplified using gene specific primers with a Strategene MX4000 thermocycler. The associated correlation coefficents for the curves are (A) 0.938 and (B) 0.935.

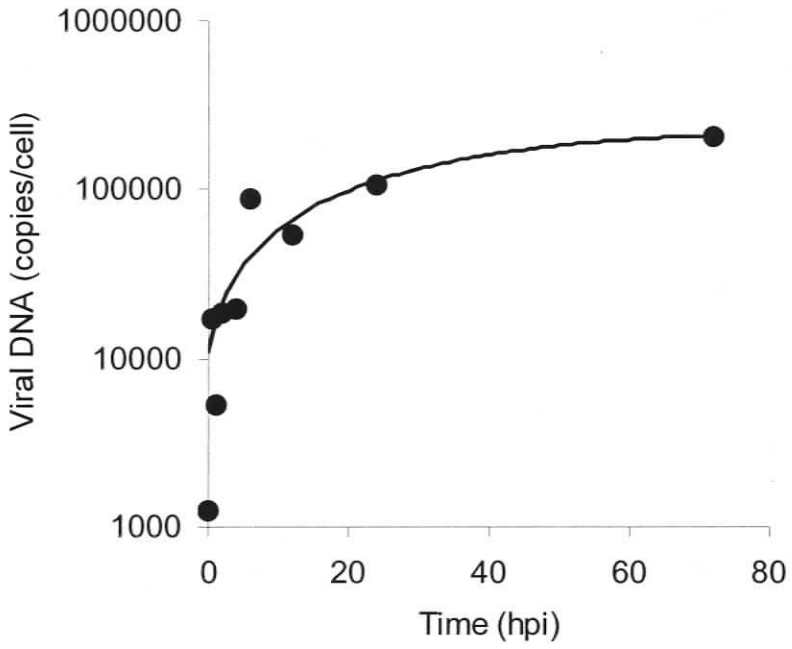


Figure 3.6. Quantitative PCR of NeabNPV-infected midgut tissue indicating the number of NeabNPV viral DNA copies per cell. PCR amplification of the NeabNPV *polh* gene was quantified and normalized against the *actin* gene, representing the number of copies of host cell genome DNA. The line fitted to the data represents a logarithmic curve with a correlation coefficient of 0.922.

3.5. Discussion

As with the lepidopteran baculoviruses, NeabNPV appears to utilize a temporal cascade of gene expression to coordinate efficient viral propagation. Our previous genome analysis (This thesis, Chapter 2) failed to identify homologues for the immediate early baculovirus genes. Putative homologues to some early-transcribed genes however, were identified in NeabNPV and transcriptional analyses indicated that early gene expression in larval midgut cells are detected within two hours of imbibing viral particles, a standard by which we may identify potential immediate early genes. The temporal gene expression pattern observed in NeabNPV is summarized in Figure 3.7.

Gene expression in lepidopteran baculoviruses is regulated at the level of transcription. For early genes this regulation is typically dependant on a small number of gene transactivators (Carson *et al.*, 1988; Guarino and Summers, 1986; Knebel-Morsdorf *et al.*, 1993; Kovacs *et al.*, 1991), whereas late genes are dependant upon the synthesis of a viral RNA polymerase (Fuchs *et al.*, 1983; Grula *et al.*, 1981; Guarino *et al.*, 1998; Huh and Weaver, 1990). Coordination of early gene expression in NeabNPV cannot be inferred, as no baculovirus transregulators have yet been identified in NeabNPV. Our observations suggest that late gene expression is temporally regulated in a manner similar to lepidopteran baculoviruses, as the RNA polymerase subunits, *lef-8* and *lef-9*, were transiently transcribed at 6 hpi and the late genes were co-ordinately transcribed after this point.

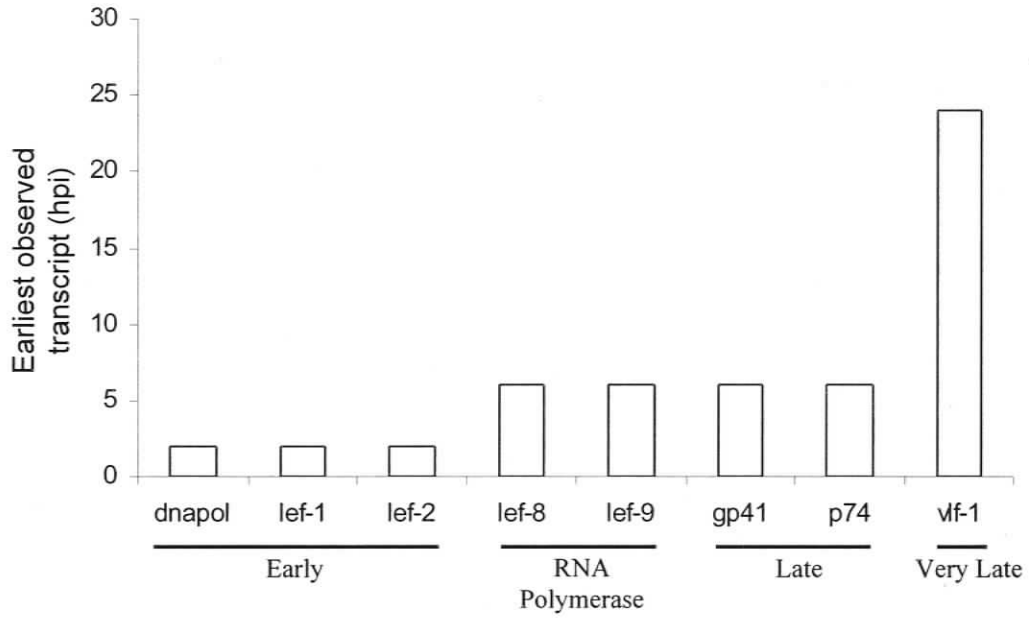


Figure 3.7. Summary of temporal gene expression patterns of NeabNPV *in vivo* infection. Illustrates the earliest times, in hours post infection (hpi), at which transcripts were observed for NeabNPV early genes (*dnapol*, *lef-1*, *lef-2*), viral-encoded RNA polymerase subunit genes (*lef-8* and *lef-9*), late genes (*gp41*, *p74*), and the gene coding for the transactivator of very late genes (*vlf-1*).

Regulation of very late expression in lepidopteran baculoviruses is initiated by the VLF-1 transactivator (McLachlin and Miller, 1994). It was expected that the very late *polh* gene transcript would be observed following that of *vlf-1*. We can infer that very late gene expression is initiated later than 24 hpi, based on *vlf-1* transcription. However, *vlf-1* transcription is later in NeabNPV than was observed in AcMNPV (McLachlin and Miller, 1994). This fact, coupled with our inability to detect transcription for the *polh* gene within 72 hpi, suggests that very late gene expression in NeabNPV may be dramatically offset compared to some lepidopteran baculoviruses. Although the AcMNPV *polh* gene transcript has been observed as early as 12 hpi (McLachlin and Miller, 1994) and is generally believed to be transcribed between 18 and 72 hpi, it may not be transcribed until as late as 96 hpi in *Perina nuda* nucleopolyhedrovirus (PenuNPV) (Chou *et al.*, 1996). As a result, there is a premise of the potential late expression of *polh* in NeabNPV.

Some baculovirus early genes, in particular the gene encoding viral DNA polymerase, have been reported as having a narrow window of expression (Tomalski *et al.*, 1988). Our study showed a similar narrow expression for the RNA polymerase subunit genes, *lef-8* and *lef-9*, but the gene transcripts, including those for the *dnapol* gene were observed until 72 hpi. One possible explanation for this observation is that, whereas *in vitro* infection may be synchronized, *in vivo* primary infection may give rise to subsequent infections within the host. The observation of late transcribed early genes in this study may not be comparable to *in vitro* studies.

We also examined the replication kinetics of NeabNPV DNA. Our initial hypothesis was that we would observe a lag period following treatment of the larvae

with NeabNPV. Lepidopteran baculovirus DNA synthesis typically lags initially followed by an acceleration by a factor of 10^3 at 8-10 hpi *in vitro* (Braunagel *et al.*, 1998; Knudson and Tinsley, 1978; Liu and Bilimoria, 1990; Rosinski *et al.*, 2002). Our DNA replication kinetics indicated that DNA synthesis increased over the first 24 hpi followed to a predicted maximum at 79 hpi.

A number of factors confound comparison of NeabNPV *in vivo* DNA replication and transcription analyses with the lepidopteran baculovirus *in vitro* studies. Firstly, different viruses and host cell species might lead to inconsistencies in the timing of gene expression and DNA replication. However, as Table 3.1 indicates, temporal gene expression of different lepidopteran baculoviruses remains relatively constant.

Additionally, non-lepidopteran baculovirus pathology differs dramatically from many lepidopteran baculoviruses. Whereas most lepidopteran baculoviruses, except the type 3 GVs, typically have a tropic range that radiates from the midgut to the hemocytes and systemic tissues, non-lepidopteran baculoviruses are midgut localized. The lepidopteran baculovirus, *Harrisina brillians* granulovirus (HabrGV) is unique, as it is midgut localized but kills its host in a similar time frame as other lepidopteran baculoviruses. Sawfly baculoviruses, such as NeabNPV, are also midgut localized (Federici, 1997), and kill their host within an average of 5-7 days (Bird and Whalen, 1953; Li, 2005; Steinhaus, 1948), a time frame that is also comparable to lepidopteran NPVs like AcMNPV. Despite the differences in pathology induced by different viruses, this study suggests that temporal gene expression and DNA replication of NeabNPV are similar to those of lepidopteran baculoviruses.

In vivo studies of viral replication and transcription patterns are more difficult than similar studies conducted *in vitro*. The lack of an *in vitro* virus propagation system for NeabNPV necessitates the use of quantitative PCR over nucleic acid hybridization, and low RNA yields (50-500 ng per larva) required the use of RT-PCR over Northern blots. Although physical and physiological barriers to infection may act as impediments to viral infection *in vivo* (*reviewed in* Wang and Granados, 2001), studies of recombinant AcMNPV infection in lavae have shown that the timing of viral DNA replication and transcription *in vivo* are similar to those observed *in vitro* (Flipsen *et al.*, 1995; Washburn *et al.*, 1995). Despite this, discrepancies in viral transcription between tissues, in lepidopteran hosts, suggests that care must be taken when comparing *in vitro* and *in vivo* systems (Pham and Sivasubramanian, 1992).

Temporal expression of the NeabNPV *pif* gene appears to be considerably different from *pif* gene expression in lepidopteran baculoviruses. The *pif* gene has been previously observed only late in infection of *S. frugiperda* and *S. littura* NPVs *in vitro* (Gutierrez *et al.*, 2004; Simon *et al.*, 2005). However, the genome sequence of NeabNPV (This thesis, Chapter 2) indicated that the *pif* gene promoter has binding site motifs for both the host zeste and CREB/ATF47 transcription factors. We observed that these motifs were coincident and almost exclusively found upstream of predicted early-expressed genes. One exception to this was the predicted late-expressed *pif* gene, which had both upstream motifs. This study indicates that the NeabNPV *pif* gene is clearly expressed early in NeabNPV-infected *N. abietis* larvae and provides further support to the hypothesized role of the host zeste and CREB/ATF47 transcription factors in early NeabNPV transcription.

This study provides the first reported analysis of transcription and viral DNA replication of a non-lepidopteran baculovirus *in vivo*. As with the lepidopteran baculoviruses, the NeabNPV utilizes a cascade of gene expression where the synthesis of a viral RNA polymerase appears to coordinate the expression of the NeabNPV late genes. We characterized a 72-hour period post-infection, where early gene expression was first observed at 2 hpi, and late gene expression at 6 hpi. Late gene expression continued until the end of DNA replication at 79 hours and very late gene expression initiated from 24 hpi. Although immediate early gene expression remains a mystery, this study provides a temporal framework by which similar *in vivo* studies can elaborate on temporal expression of NeabNPV.

Chapter 4. Sequence and transcriptional analysis of two potential zinc finger proteins of the *Neodiprion abietis* nucleopolyhedrovirus.

4.1. Abstract

The mechanism of early gene transcription in baculoviruses isolated from sawflies (order *Hymenoptera*) is poorly characterized, as no known homologues of lepidopteran baculovirus immediate early genes have been identified. We present DNA sequence and transcriptional analyses of two novel genes, *neab24* and *neab52*, of the Balsam fir sawfly, *Neodiprion abietis*, nucleopolyhedrovirus (NeabNPV). Transcription of *neab24* was observed within 0.5 hours post infection (hpi), suggesting that it is possibly expressed as an immediate-early gene, and *neab52* was expressed early by 2 hpi. Both predicted gene products contain zinc finger motifs. The potential of these motifs as regulatory elements is discussed.

4.2. Introduction

Baculoviruses are a family of mostly insect-specific enveloped viruses with large double-stranded genomes and complex gene expression. The best-characterized baculoviruses are those that infect lepidopteran hosts. To expand our understanding of the non-lepidopteran baculoviruses we have previously reported the genome sequence of the hymenopteran baculovirus *Neodiprion abietis* nucleopolyhedrovirus (NeabNPV) (This thesis, Chapter 2). We also conducted a study to map the temporal transcription cascade of NeabNPV gene expression (This thesis, Chapter 3).

In lepidopteran baculoviruses, gene expression is temporally coordinated into distinct stages: immediate early, early, late and very late. Gene expression is primarily regulated at the level of transcription, with immediate early genes acting as transactivators for delayed early genes (Glocker *et al.*, 1992). The delayed early genes include a virally encoded RNA polymerase, composed of the products of *lef-4*, *lef-8*, *lef-9* and *p47*, that promotes late gene expression (Guarino *et al.*, 1998). Among the late genes very late factor (*vlf-1*), encodes a transactivator that subsequently promotes very late gene expression (McLachlin and Miller, 1994; Mistretta and Guarino, 2005; Todd *et al.*, 1996; Yang and Miller, 1999).

Four non-lepidopteran baculovirus genomes have been fully sequenced, including the nucleopolyhedrovirus (NPV) of the mosquito (order: *Diptera*) *Culex nigripalpus* (CuniNPV) and the sawfly (order: *Hymenopteran*) NPVs of *Neodiprion sertifer* (NeseNPV), *Neodiprion lecontei* (NeleNPV) and *N. abietis* (Afonso *et al.*, 2001; Garcia-Maruniak *et al.*, 2004; Lauzon *et al.*, 2004, This thesis, Chapter 2). Although the non-lepidopteran NPVs share conserved homologues of early, late and very late

genes with the lepidopteran NPVs, no homologues of lepidopteran conserved immediate early genes have yet been identified.

Our genome sequence analysis of NeabNPV identified a number of putative ORFs with characteristics that may be indicative of immediate early and early genes (This thesis, Chapter 2). As indicated in Table 4.1, immediate early gene products of lepidopteran baculoviruses typically have characteristic amino acid motifs that confer the ability to bind DNA and transactivate downstream genes. Two uncharacterized NeabNPV genes, *neab24* and *neab52*, encode proteins with putative zinc-finger motifs. Zinc-finger motifs have been reported to facilitate DNA binding and were identified in AcMNPV transactivator PE-38 (Krappa and Knebel-Morsdorf, 1991). Furthermore, our genome analysis revealed that a Zeste binding site was enriched in the sequences within 160 bp upstream of early baculovirus genes (This thesis, Chapter 2). The *neab52* gene in NeabNPV also has a predicted upstream Zeste-binding site.

To examine the potential roles each of these genes, *neab24* and *neab52*, in NeabNPV infection, we analyzed the coding and upstream sequences of these genes. We have previously established that early-expressed NeabNPV genes are observed from 2 hours post infection (This thesis, Chapter 3). We further examined the temporal expression of *neab24* and *neab52*. Our analysis suggests that both genes appear to be early expressed. Furthermore, *neab24* appears to be a candidate immediate early gene in NeabNPV.

Table 4.1. Amino acid motifs and earliest observed expression of immediate early baculovirus genes.

Baculovirus Gene	Earliest expressed [a]	Amino acid motifs	References
IE-1/0	1 hpi (AcMNPV, SpexNPV) 0 hpi (OpMNPV)	N-terminal acidic domain. C-terminal DNA-binding domain.	(Chisholm and Henner, 1988; Kovacs *et al.*, 1992; Kovacs *et al.*, 1991; Theilmann and Stewart, 1991; van Strien *et al.*, 2000)
IE-2	0.5 hpi (OpMNPV)	N-terminal highly basic domain (OpMNPV). RING finger (OpMNPV). Ser-Thr rich region Pro-rich region Poly-Glu tract	(Prikhod'ko and Miller, 1998; Theilmann and Stewart, 1992;) (Carson *et al.*, 1988)
PE38	1 hpi	Zinc finger Leucine zipper	(Krappa and Knebel-Morsdorf, 1991)
ME53	1 hpi	N-terminal zinc finger.	(Knebel-Morsdorf *et al.*, 1993)

[a] hpi = hours post infection

4.3. Methods and Materials

Virus Source. NeabNPV was isolated in 1997 from N. abietis larvae collected between Stephenville (latitude 48°32' N, longitude 58°33' W) and Corner Brook. The virus was purified as described by Moreau *et al.* (2005) and as described in the 'Methods and Materials' section of chapter 2.

Larval infection. Balsam Fir sawfly larvae (*N. abietis*) were collected and infected with NeabNPV as described in the 'Methods and Materials' section of chapter 3.

Viral RNA isolation. Total RNA was extracted from NeabNPV-infected Balsam Fir sawfly larvae (*N. abietis*) as described in the 'Methods and Materials' section of chapter 3.

RNA Reverse Transcription PCR. The RNA samples were treated with DNAse and RT-PCR was performed as described in the 'Methods and Materials' section of chapter 3.

Two primer sets were used in this experiment: the predicted *neab24 gene* was amplified with the primers 5'-GCCCACGGCGTAGTGTGTTGT-3' and 5'-GGGCCATGATATGTCGGCAA-3', and the predicted *neab52 gene* was amplified with the primers 5'-CGCGTTTTGAACGTCATCGTCA-3' and 5'-ACGGTTTTTCTTTCGTATGCGTTCT-3'. As a control, the host *28S* rRNA gene was amplified with the primers 5'- AAAGATCGAATGGGGAGATTCATC-3' and 5'- CGTCCTACTAGGGGAGAAGTGCAC-3'. A positive control (NeabNPV

genomic DNA), a negative control (cDNA from uninfected larva), and a host control (28S rRNA gene-specific primers) were included in each PCR set. The DNA templates were amplified with gene specific primers and Platinum *Taq* polymerase (Invitrogen), using manufacturer's protocol. Temperature cycling was 95°C (9 mins.) followed by 45 cycles of 95°C (30 secs.), 55°C (60 secs.), and 72°C (90 secs). All products were resolved by gel electrophoresis in 2% agarose and stained with SYBR Green.

Sequence Analysis. The gene sequences for *neab24* and *neab52* were translated and their molecular weights and pI were predicted using the Expasy server (http://us.expasy.org/tools/pi_tool.html). The translated sequence was compared with those on GenBank using BLAST (Altschul *et al.*, 1990) to identify potential sequence orthologues. For the predicted *neab24* gene product, a phylogenetic tree was generated by aligning predicted protein sequences of related genes using CLUSTALW (Thompson *et al.*, 1994), constructing a tree using maximum parsimony with PAUP 4.0 b10 (Swofford,D.L.), and bootstrapping the resultant tree (1000 replicates). Multiple alignments were illustrated using BioEdit 5.0.9 (Hall, 1999).

4.4. Results

Sequence analysis of the *neab24* gene and its predicted product. The *neab24* gene is located 22,750 bp downstream from the predicted translation start site of the *polh* gene. The *neab24* ORF is 1,464 bp and potentially encodes an acidic protein of 56.9 kDa (pI=4.98). This predicted gene product encodes a putative C_4HC_4 zinc finger with a consensus sequence C-X2-C-X(9-39)-C-X(1-3)-H-X(2-3)-C/H-X2-C-X(4-48)-C-X2-C (Figure 4.1), which resembles the specialized RING finger family of zinc finger proteins (Borden and Freemont, 1996). Although neab24 shares 50% amino acid sequence identity with the NeleNPV predicted nele8 protein (AAQ99052.1), this conservation does not extend to NeseNPV, where the most similar predicted protein nese11 (YP_025118.1), exhibits only 16% global amino acid identity. In addition to nese11, the NeseNPV genome encodes three additional predicted RING-finger proteins, including nese9 (YP_025116.1), nese18 (YP_025125.1), and nese19 (YP_025126.1). Although the RING-finger proteins of NeseNPV appear to have amino acid sequence distinct from the neab24 and nele8 (Figure 4.2B), all of these predicted proteins have a conserved RING finger domain (Figure 4.2A), and thus they are likely to represent a protein family in the sawfly NPVs. BLAST analyses indicated that this proposed protein family does not have homologues in other baculoviruses. Outside of the sawfly baculoviruses, the greatest amino acid sequence similarity with a characterized protein is shared between neab24 and the 2557-2762 amino acid region of the *Plasmodium falciparum* normocyte-binding protein (AAL38218.2; 24% identity, 48% similarity).

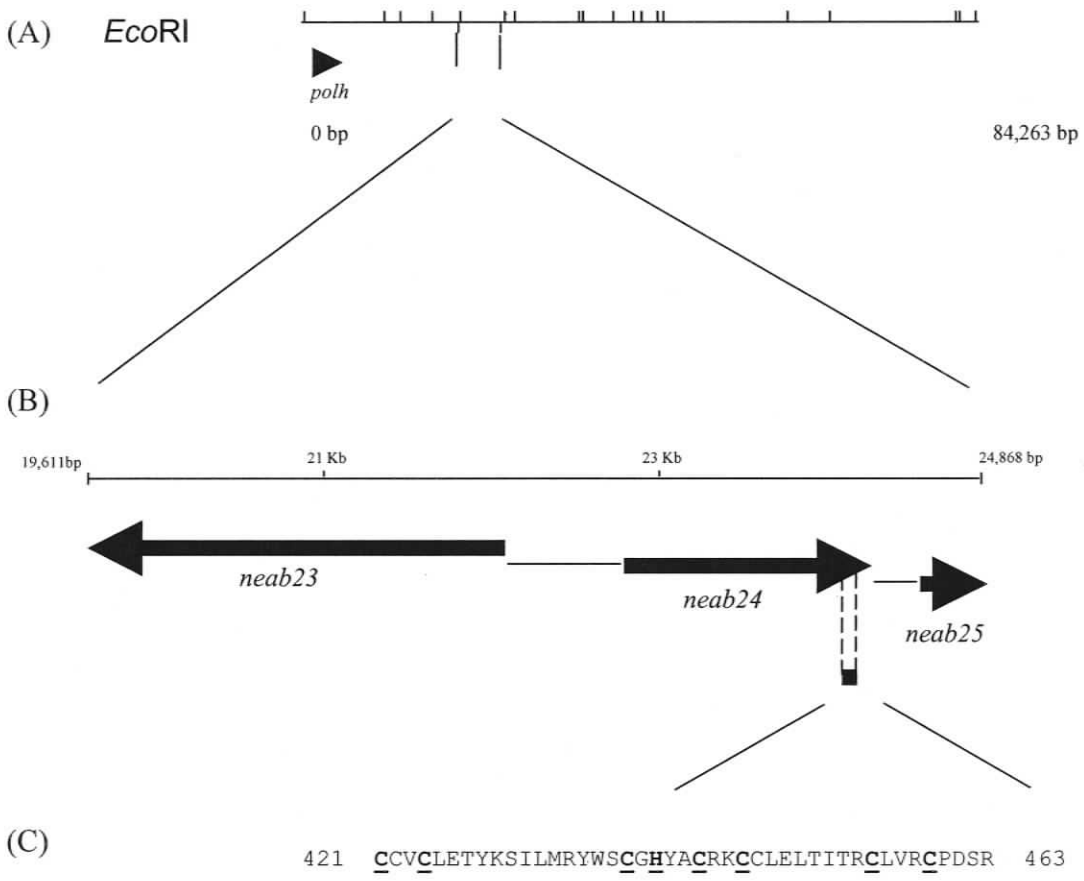
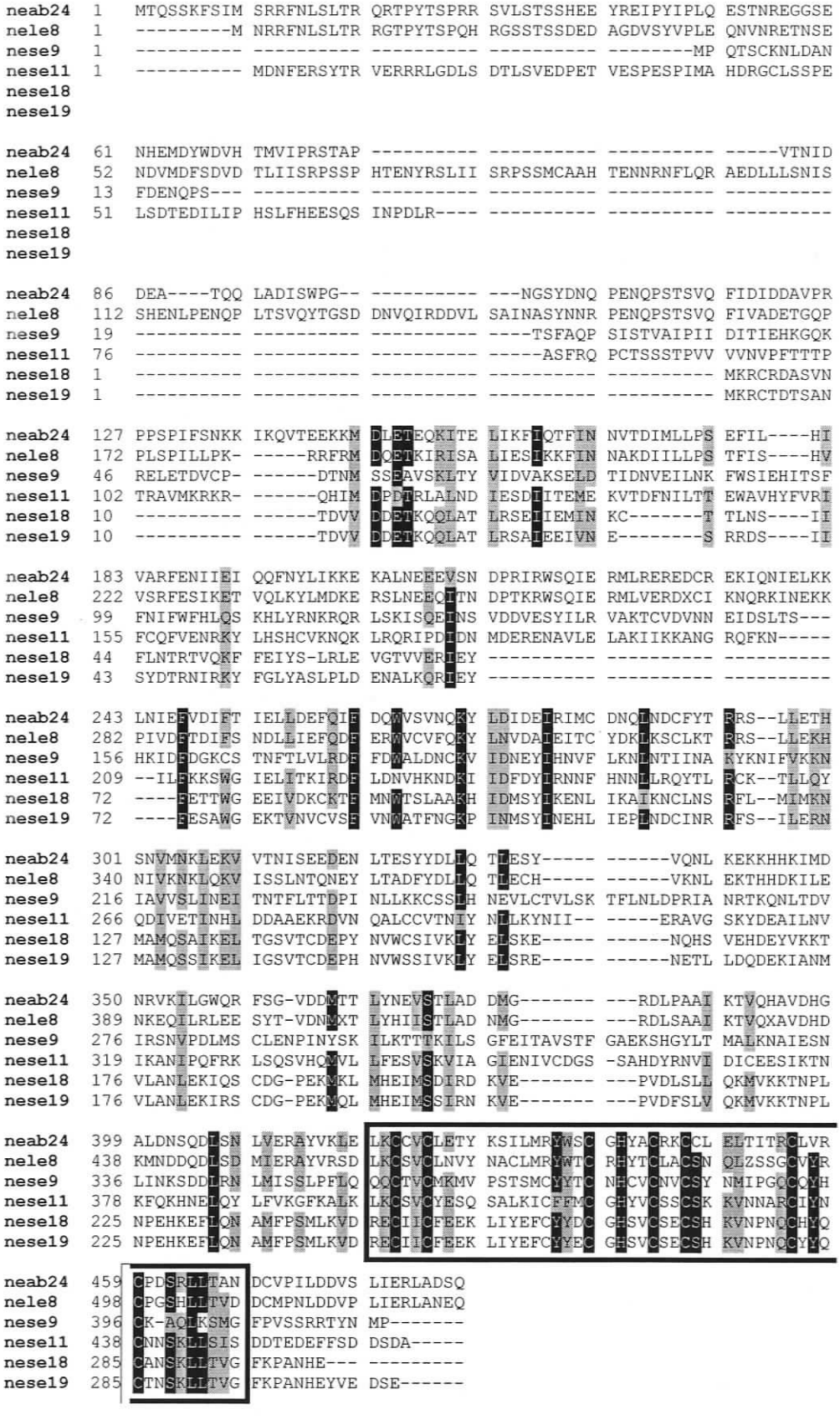


Figure 4.1. Location of the *neab24* gene on a linear map of the circular NeabNPV genome. (A) *Eco*RI restriction map of the NeabNPV genome relative to the *polh* locus (B) Magnification of the NeabNPV genome locus between 19.6 and 24.9 Kb. The distance from the *polh* translation start site is indicated above the line and the orientation of *neab23*, *neab24* and *neab25* ORFs are indicated as arrows. The line below *neab24* indicates the position of the predicted RING finger domain. (C) The amino acid sequence of the predicted RING finger domain. The cysteine and histidine are bold and underlined.

A.


neab24 1 MTQSSKFSIM SRRFNLSLTR QRTPYTSPRR SVLSTSSHEE YREIPYIPLQ ESTNREGGSE
nele8 1 ---------M NRRFNLSLTR RGTPYTSPQH RGSSTSSDED AGDVSYVPLE QNVNRETNSE
nese9 1 ---------- ---------- ---------- ---------- --------MP QTSCKNLDAN
nese11 1 ---------- MDNFERSYTR VERRRLGDLS DTLSVEDPET VESPESPIMA HDRGCLSSPE
nese18
nese19
neab24 61 NHEMDYWDVH TMVIPRSTAP ---------- ---------- ---------- -----VTNID
nele8 52 NDVMDFSDVD TLIISRPSSP HTENYRSLII SRPSSMCAAH TENNRNFLQR AEDLLLSNIS
nese9 13 FDENQPS--- ---------- ---------- ---------- ---------- ----------
nese11 51 LSDTEDILIP HSLFHEESQS INPDLR---- ---------- ---------- ----------
nese18
nese19
neab24 86 DEA----TQQ LADISWPG-- ---------- ---NGSYDNQ PENQPSTSVQ FIDIDDAVPR
nele8 112 SHENLPENQP LTSVQYTGSD DNVQIRDDVL SAINASYNNR PENQPSTSVQ FIVADETGQP
nese9 19 ---------- ---------- ---------- ----TSFAQP SISTVAIPII DITIEHKGQK
nese11 76 ---------- ---------- ---------- -----ASFRQ PCTSSSTPVV VVNVPFTTTP
nese18 1 ---------- ---------- ---------- ---------- ---------- MKRCRDASVN
nese19 1 ---------- ---------- ---------- ---------- ---------- MKRCTDTSAN
neab24 127 PPSPIFSNKK IKQVTEEKKM DLETEQKITE LIKFIQTFIN NVTDIMLLPS EFIL----HI
nele8 172 PLSPILLPK- -----RRFRM DQETKIRISA LIESIKKFIN NAKDIILLPS TFIS----HV
nese9 46 RELETDVCP- ------DTNM SSEAVSKLTY VIDVAKSELD TIDNVEILNK FWSIEHITSF
nese11 102 TRAVMKRKR- ------QHIM DPDTRLALND IESDIITEME KVTDFNILTT EWAVHYFVRI
nese18 10 ---------- ------TDVV DDETKQQLAT LRSEIIEMIN KC-------T TLNS----II
nese19 10 ---------- ------TDVV DDETKQQLAT LRSAIEEIVN E--------S RRDS----II
neab24 183 VARFENIIEI QQFNYLIKKE KALNEEEVSN DPRIRWSQIE RMLREREDCR EKIQNIELKK
nele8 222 VSRFESIKET VQLKYLMDKE RSLNEEQITN DPTKRWSQIE RMLVERDXCI KNQRKINEKK
nese9 99 FNIFWFHLQS KHLYRNKRQR LSKISQEINS VDDVESYILR VAKTCVDVNN EIDSLTS---
nese11 155 FCQFVENRKY LHSHCVKNQK LRQRIPDIDN MDERENAVLE LAKIIKKANG RQFKN-----
nese18 44 FLNTRTVQKF FEIYS-LRLE VGTVVERIEY ---------- ---------- ----------
nese19 43 SYDTRNIRKY FGLYASLPLD ENALKQRIEY ---------- ---------- ----------
neab24 243 LNIEFVDIFT IELLDEFQIE DQWVSVNQKY LDIDEIRIMC DNQLNDCFYT FRS--LLETH
nele8 282 PIVDFTDIFS NDLLIEFQDE ERWVCVFQKY LNVDAIEITC YDKLKSCLKT FRS--LLEKH
nese9 156 HKIDFDGKCS TNFTLVLRDE FDWALDNCKV IDNEYIHNVF LKNLNTIINA KYKNIFVKKN
nese11 209 --ILFKKSWG IELITKIRDE LDNVHKNDKI IDFDYIRNNF HNNLLRQYTL FCK--TLLQY
nese18 72 ----FETTWG EEIVDKCKTE MNWTSLAAKH IDMSYIKENL IKAIKNCLNS FFL--MIMKN
nese19 72 ----FESAWG EKTVNVCVSE VNWATFNGKP INMSYINEHL IEPLNDCINR FFS--ILERN
neab24 301 SNVMNKLEKV VTNISEEDEN LTESYYDLLQ TLESY----- ------VQNL KEKKHHKIMD
nele8 340 NIVKNKLQKV ISSLNTQNEY LTADFYDLLQ TLECH----- ------VKNL EKTHHDKILE
nese9 216 IAVVSLINEI TNTFLTTDPI NLLKKCSSLH NEVLCTVLSK TFLNLDPRIA NRTKQNLTDV
nese11 266 QDIVETINHL DDAAEKRDVN QALCCVTNIY NLLKYNII-- -----ERAVG SKYDEAILNV
nese18 127 MAMQSAIKEL TGSVTCDEPY NVWCSIVKEY ELSKE----- ------NQHS VEHDEYVKKT
nese19 127 MAMQSSIKEL IGSVTCDEPH NVWSSIVKEY ELSRE----- ------NETL LDQDEKIANM
neab24 350 NRVKILGWQR FSG-VDDMTT LYNEVSTLAD DMG------- ---RDLPAAI KTVQHAVDHG
nele8 389 NKEQILRLEE SYT-VDNMXT LYHIISTLAD NMG------- ---RDLSAAI KTVQXAVDHD
nese9 276 IRSNVPDLMS CLENPINYSK ILKTTTKILS GFEITAVSTF GAEKSHGYLT MALKNAIESN
nese11 319 IKANIPQFRK LSQSVHQVVL LFESVSKVIA GIENIVCDGS -SAHDYRNVI DICEESIKTN
nese18 176 VLANLEKIQS CDG-PEKVKL MHEIMSDIRD KVE------- ---PVDLSLL QKMVKKTNPL
nese19 176 VLANLEKIRS CDG-PEKMQL MHEIMSSIRN KVE------- ---PVDFSLV QKMVKKTNPL
neab24 399 ALDNSQDLSN LVERAYVKLE LKCCVCLETY KSILMRYWSC GHYACRKCCL ELTITRCLVR
nele8 438 KMNDDQDLSD MIERAYVRSD LKCSVCLNVY NACLMRYWTC RHYTCLACSN QLZSSGCVVR
nese9 336 LINKSDDLRN LMISSLPFLQ QQCTVCMKMV PSTSMCYYTC NHCVCNVCSY NMIPGCCQYH
nese11 378 KFQKHNELQY LFVKGFKALK LKCSVCYESQ SALKICFFMC GHYVCSSCSK KVNNARCIYN
nese18 225 NPEHKEFLQN AMFPSMLKVD RECIICFEEK LIYEFCYYDC GHSVCSECSH KVNPNQCHYQ
nese19 225 NPEHKEFLQN AMFPSMLKVD RECIICFEEK LIYEFCYYEC GHSVCSECSH KVNPNQCYYQ
neab24 459 CPDSRLLTAN DCVPILDDVS LIERLADSQ
nele8 498 CPGSHLLTVD DCMPNLDDVP LIERLANEQ
nese9 396 CK-AQLKSMG FPVSSRRTYN MP-------
nese11 438 CNNSKLLSIS DDTEDEFFSD DSDA-----
nese18 285 CANSKLLTVG FKPANHE--- ---------
nese19 285 CTNSKLLTVG FKPANHEYVE DSE------

B.

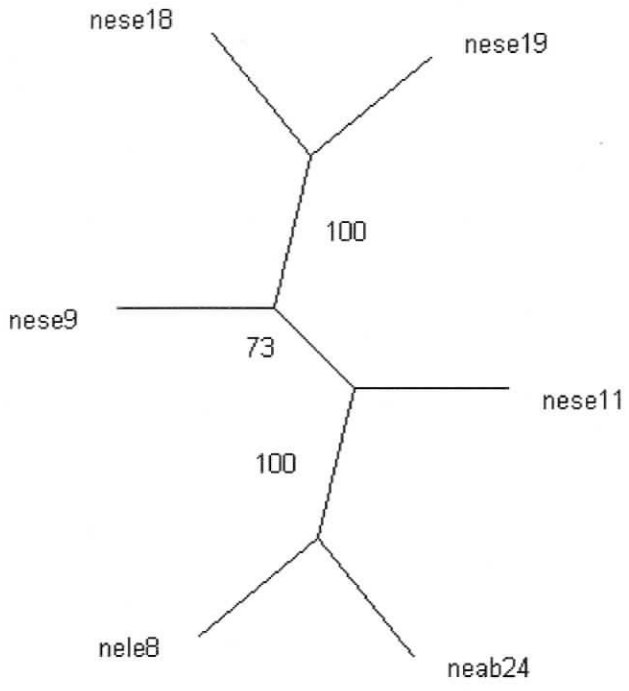


Figure 4.2. Amino acid sequence alignment of NeabNPV, NeseNPV, and NeleNPV RING finger proteins and phylogenetic analysis. (A) Multiple amino acid alignment of the RING finger motif for the predicted proteins NeabNPV neab24, NeleNPV nele8, and NeseNPV nese9, nese11, nese18, and nese19. White text on black background indicates sequence identity whereas black text on grey indicates sequence similarity between aligned amino acids. The RING finger domain is outlined. (B) Maximum parsimony phylogeny of the aligned complete amino acid sequences of each protein. The numbers beside the branches indicate the percent of bootstrap replicates (out of 1000) that support the clade.

Sequence analysis of the *neab52* gene and its predicted product. The *neab52* gene is located 47, 917 bp downstream from the predicted translation start site of the *polh* gene. The *neab52* ORF is 669 bp and potentially encodes a basic protein of 26.8 kDa in size (pI=9.38). This predicted gene product encodes four putative CCHH zinc finger domains with the motif sequence C-X_{1-2}-C-X_{8-12}-H-X_{3-4}-H (Figure 4.3). Based on BLAST analysis of the predicted neab52 amino acid sequence, this protein shares amino acid sequence similarity with predicted proteins of *N. lecontei* (nele49; AAQ99097.1) and *N. sertifer* (nese52; YP_025159.1). Sequence alignment suggests that neab52 has two domains (Figure 4.4). The first domain, between amino acids 1-88, encodes the four zinc finger motifs. Sequence alignment with the top BLAST hits indicated that the first domain was conserved in zebrafish (*Danio rerio*) predicted p135 (XP_694059.1), cow (*Bos taurus*) predicted p95 (XP_591561.2), sea urchin (*Stronglyocentrotus purpuratus*) predicted p569 (XP_790426.1), and human (*Homo sapiens*; CAD39111.1) (Figure 4.5). The second domain, between amino acids 118-182, does not appear to have sequence identity with any other baculovirus proteins.

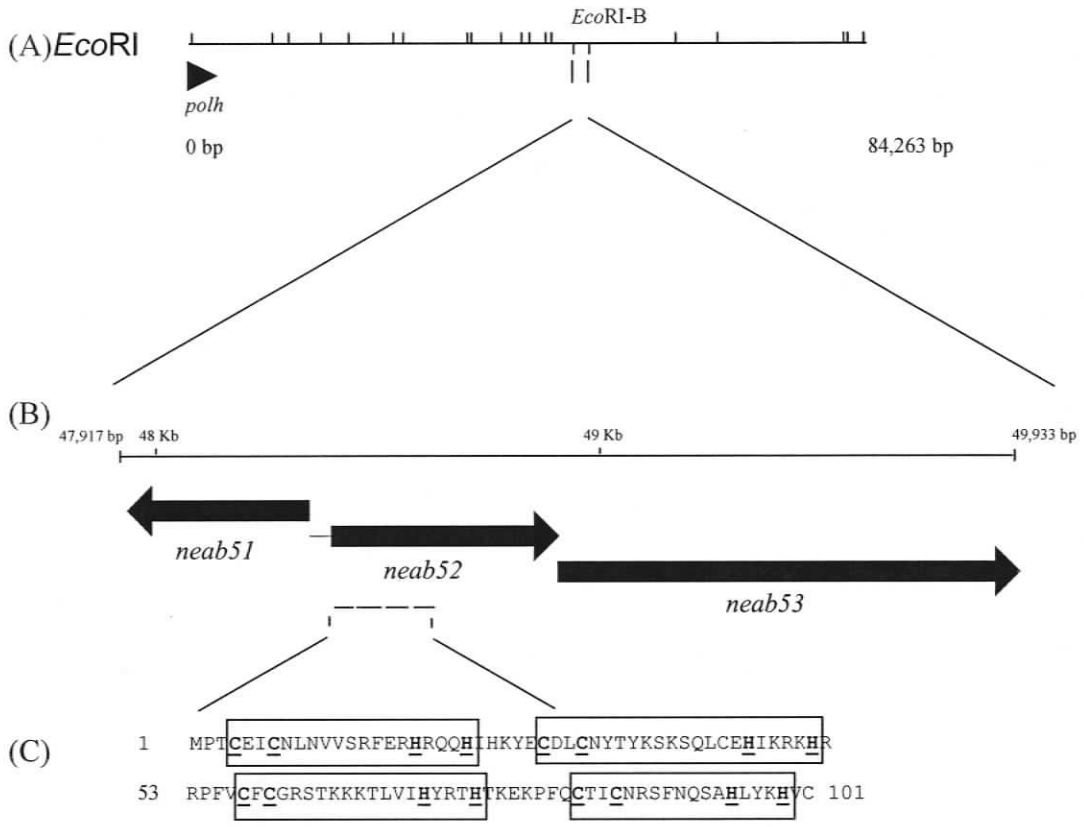


Figure 4.3. Location of the *neab52* gene on a linear map of the circular NeabNPV genome. (A) *Eco*RI restriction map of the NeabNPV genome relative to the *polh* locus (B) Magnification of the NeabNPV genome locus between 47.9 and 49.9 Kb. The distance from the *polh* translation start site is indicated above the line and the orientation of *neab51*, *neab52* and *neab53* ORFs are indicated as arrows. Dotted line below *neab52* indicates the positions of the four predicted zinc-finger motifs. (C) The amino acid sequence of the N-terminus of the *neab52* predicted gene product indicating the four predicted zinc-finger motifs outlined with boxes. The cysteine and histidine are bold and underlined.

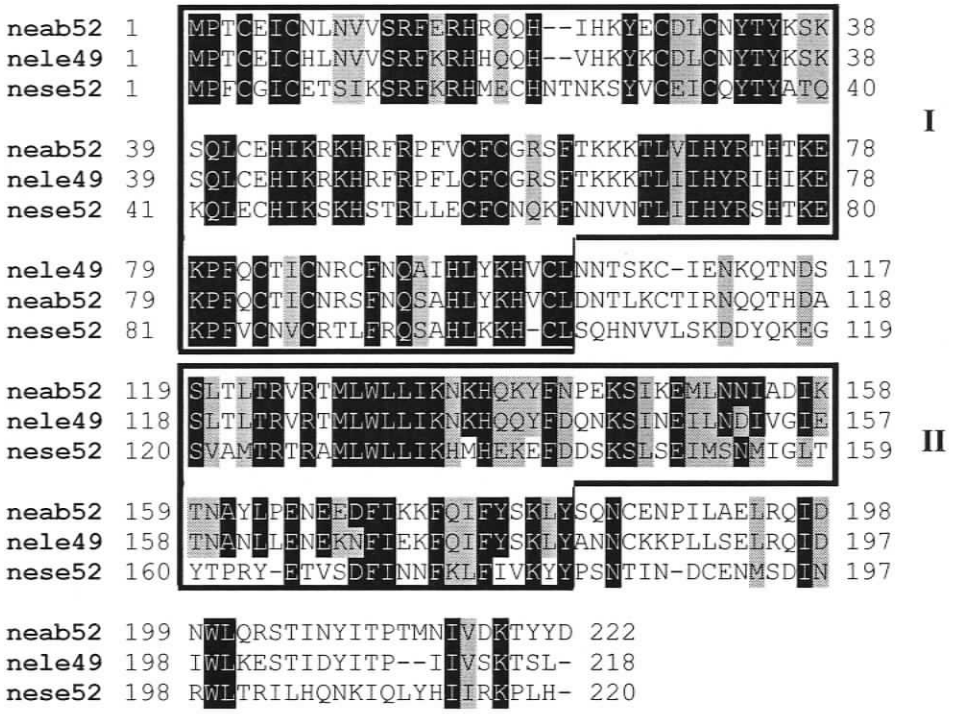


Figure 4.4. Sequence alignment of the predicted protein homologues of neab52, nele49 from *Neodiprion lecontei* nucleopolyhedrovirus and nese52 from *Neodiprion sertifer* nucleopolyhedrovirus. White text on grey background highlights identical amino acids whereas black text on grey background indicates similar amino acids. Two predicted conserved regions have been outlined and labelled.

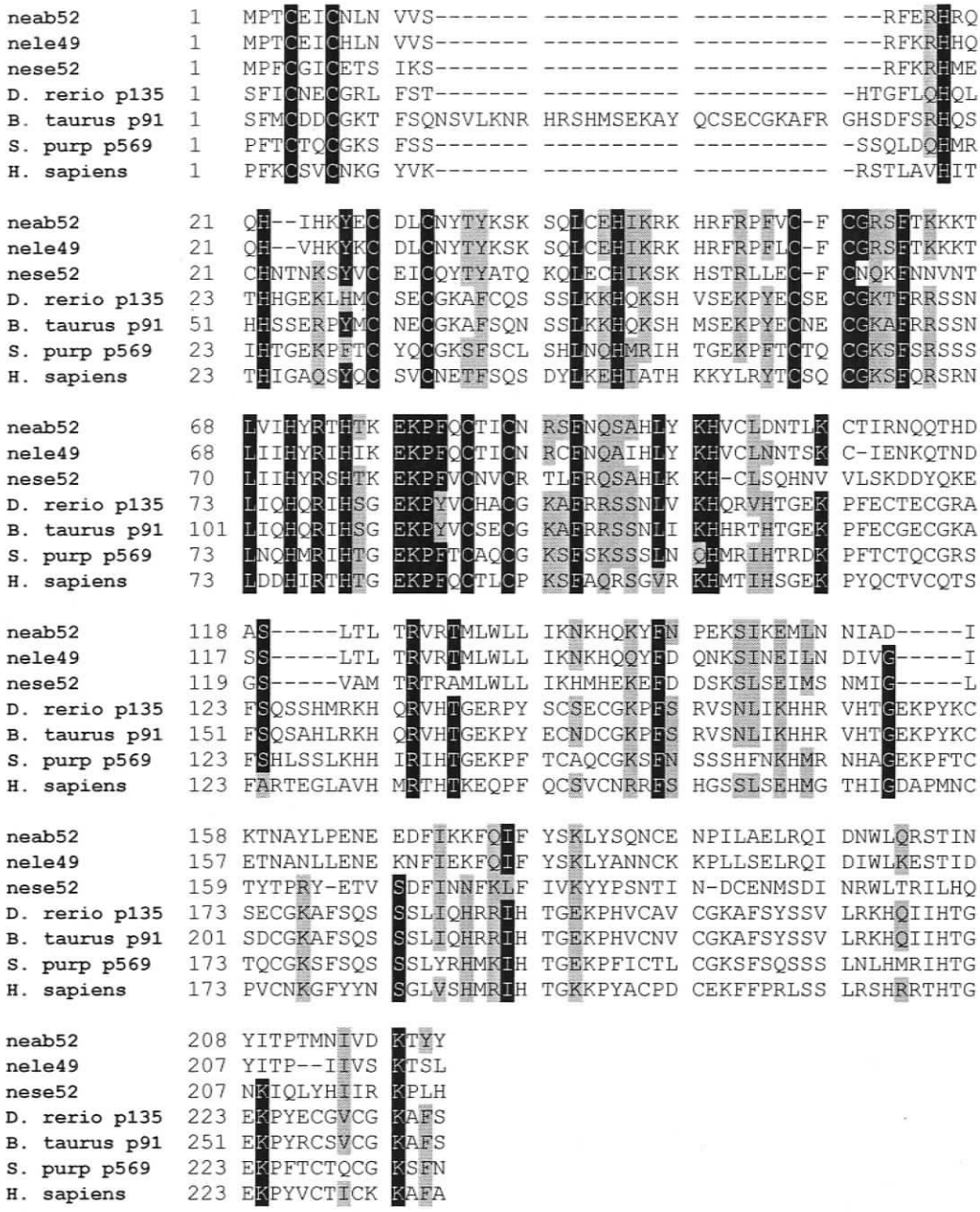

neab52 1 MPTCEICNLN VVS------- ---------- ---------- ---RFERHRQ
nele49 1 MPTCEICHLN VVS------- ---------- ---------- ---RFKRHHQ
nese52 1 MPFCGICETS IKS------- ---------- ---------- ---RFKRHME
D. rerio p135 1 SFICNECGRL FST------- ---------- ---------- -HTGFLQHQL
B. taurus p91 1 SFMCDDCGKT FSQNSVLKNR HRSHMSEKAY QCSECGKAFR GHSDFSRHQS
S. purp p569 1 PFTCTQCGKS FSS------- ---------- ---------- -SSQLDQHMR
H. sapiens 1 PFKCSVCNKG YVK------- ---------- ---------- -RSTLAVHIT
neab52 21 QH--IHKYEC DLCNYTYKSK SQLCEHIKRK HRFRPFVC-F CGRSFTKKKT
nele49 21 QH--VHKYKC DLCNYTYKSK SQLCEHIKRK HRFRPFLC-F CGRSFTKKKT
nese52 21 CHNTNKSYVC EICQYTYATQ KQLECHIKSK HSTRLLEC-F CNQKFNNVNT
D. rerio p135 23 THHGEKLHMC SECGKAFCQS SSLKKHQKSH VSEKPYECSE CGKTFRRSSN
B. taurus p91 51 HHSSERPYMC NECGKAFSQN SSLKKHQKSH MSEKPYECNE CGKAFRRSSN
S. purp p569 23 IHTGEKPFTC YQCGKSFSCL SHLNQHMRIH TGEKPFTCTQ CGKSFSRSSS
H. sapiens 23 THIGAQSYQC SVCNETFSQS DYLKEHIATH KKYLRYTCSQ CGKSFQRSRN
neab52 68 LVIHYRTHTK EKPFQCTICN RSFNQSAHLY KHVCLDNTLK CTIRNQQTHD
nele49 68 LIIHYRIHIK EKPFQCTICN RCFNQAIHLY KHVCLNNTSK C-IENKQTND
nese52 70 LIIHYRSHTK EKPFVCNVCR TLFRQSAHLK KH-CLSQHNV VLSKDDYQKE
D. rerio p135 73 LIQHQRIHSG EKPYVCHACG KAFRRSSNLV KHQRVHTGEK PFECTECGRA
B. taurus p91 101 LIQHQRIHSG EKPYVCSECG KAFRRSSNLI KHHRTHTGEK PFECGECGKA
S. purp p569 73 LNQHMRIHTG EKPFTCAQCG KSFSKSSSLN QHMRIHTRDK PFTCTQCGRS
H. sapiens 73 LDDHIRTHTG EKPFQCTLCP KSFAQRSGVR KHMTIHSGEK PYQCTVCQTS
neab52 118 AS-----LTL TRVRTMLWLL IKNKHQKYFN PEKSIKEMLN NIAD-----I
nele49 117 SS-----LTL TRVRTMLWLL IKNKHQQYFD QNKSINEILN DIVG-----I
nese52 119 GS-----VAM TRTRAMLWLL IKHMHEKEFD DSKSLSEIMS NMIG-----L
D. rerio p135 123 FSQSSHMRKH QRVHTGERPY SCSECGKPFS RVSNLIKHHR VHTGEKPYKC
B. taurus p91 151 FSQSAHLRKH QRVHTGEKPY ECNDCGKPFS RVSNLIKHHR VHTGEKPYKC
S. purp p569 123 FSHLSSLKHH IRIHTGEKPF TCAQCGKSFN SSSHFNKHMR NHAGEKPFTC
H. sapiens 123 FARTEGLAVH MRTHTKEQPF QCSVCNRRFS HGSSLSEHMG THIGDAPMNC
neab52 158 KTNAYLPENE EDFIKKFQIF YSKLYSQNCE NPILAELRQI DNWLQRSTIN
nele49 157 ETNANLLENE KNFIEKFQIF YSKLYANNCK KPLLSELRQI DIWLKESTID
nese52 159 TYTPRY-ETV SDFINNFKLF IVKYYPSNTI N-DCENMSDI NRWLTRILHQ
D. rerio p135 173 SECGKAFSQS SSLIQHRRIH TGEKPHVCAV CGKAFSYSSV LRKHQIIHTG
B. taurus p91 201 SDCGKAFSQS SSLIQHRRIH TGEKPHVCNV CGKAFSYSSV LRKHQIIHTG
S. purp p569 173 TQCGKSFSQS SSLYRHMKIH TGEKPFICTL CGKSFSQSSS LNLHMRIHTG
H. sapiens 173 PVCNKGFYYN SGLVSHMRIH TGKKPYACPD CEKFFPRLSS LRSHRRTHTG
neab52 208 YITPTMNIVD KTYY
nele49 207 YITP--IIVS KTSL
nese52 207 NKIQLYHIIR KPLH
D. rerio p135 223 EKPYECGVCG KAFS
B. taurus p91 251 EKPYRCSVCG KAFS
S. purp p569 223 EKPFTCTQCG KSFN
H. sapiens 223 EKPYVCTICK KAFA

Figure 4.5. Multiple sequence alignment of putative homologues of NeabNPV neab52. Identical amino acid sequences are illustrated as white text on black background and similar amino acid sequences are illustrated as black text on gray background. neab52, accession no. ABC74926; nele49, accession no. YP_025246.1; nese52, accession no. AAQ96429.1; *D. rerio* p135, accession no. XP_694059.1; *B. taurus* p91, accession no. XP_591561.2; *S. purpuratus* p569, accession no.XP_790426.1; *H. sapiens*, accession no. CAD39111.1.

Promoter sequence and transcriptional analysis of *neab24* and *neab52*. The sequences 500 bp upstream of the predicted ORFs were queried using SIGSCAN against the TRANSFAC database with a query limited to insects. For both genes, this analysis only revealed TATA-related and Zeste binding sites on the coding strand. For the *neab24* ORF a proximal putative TATA box was identified 23 bp upstream of the predicted translation start site. Distal TATA box sites were detected at –360 and –462 bp. A distal putative Zeste binding motif was identified at –345 bp (Figure 4.6). The *neab52* ORF had a proximal putative TATA box at –75 bp and Zeste-binding site at –134, relative to the predicted translation start site. Distal putative TATA box sites were identified at –169, -283, -351 and –473 bp (Figure 4.7).

RT-PCR analysis of the *neab24* and *neab52* gene transcription revealed that these genes are transcribed from 0.5 or 2 hpi, respectively. The only exception to this was an amplicon was not observed for *neab52* at 12 hpi. Based on our previous work, mapping the temporal expression of NeabNPV (This Thesis, Chapter 3), both of these genes are early-expressed as they are both observed by 2 hpi. Given that *neab24* is observed as early as 0.5 hpi, it may also be considered as a candidate immediate early gene.

(A)

TATA

-500 TACGTTGGAA CTGTGGTTTA CCGTACACGG ATTCGTTTAT ATACGAACCA GTAACGTACG ATAACTAGAA

-430 TCTAAGATTA CAGGTATCTT GAAACACTTG ATATCAAAAT TAATTAACCG ACCGTAAATC AATCATAGGT

TATA Zeste

-360 ATACAGGACA AATCCGAGTG ATTATCACAC TTGTTTGTCA CACCTAATCG TTGATCTTTG CGTGGTCACG

-290 TAGTCGTATC AATCAAGCAG ATAATACTTT AATCAATCAT GTTGACACTA CGACCAGTCA TGATATCTGC

-220 CAACCGGTTG TTACAGACAT TGATCGATCG TGTCGTTTGT CACGTAGGCC AATCAATCAA ATAGATTACA

-150 TCGTCGATAA ATAATTTTTA ATCAAGACGG TGTTATCAGA GTTTGTCAGC CGTGCTGTTT GTCACGTAGA

TATA

-80 CCTGAATGAT AACACGTATC GACTGATAGT TGTTATCAAA AAATGATAAC GTTCGATATA AAGAGCCGGA

-10 TCACCGGCAC **ATG**ACACAGT CTTC

(B)

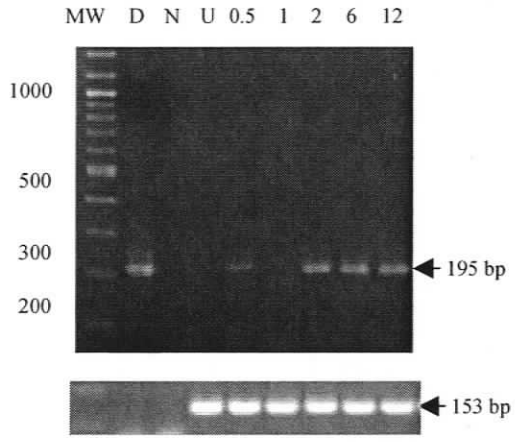


Figure 4.6. Promoter sequence and RT-PCR analysis of the *neab24* gene. (A) Promoter sequence analysis of the sequence 500 bp upstream of the predicted translation start site for the *neab24* gene. Putative TATA and Zeste transcription factor binding motifs are underlined and labelled and the predicted translation start site is in bold. (B) RT-PCR was performed from RNA extracted from uninfected and NeabNPV-infected larve. The expected size of the RT-PCR product is indicated to the right of the panel. Lane D) represents amplification from NeabNPV DNA (positive control) and; Lane N) amplification from no template; and Lane U) amplification from uninfected larvae (negative control). The remaining lanes are amplicons from RNA extracted from larvae 0.5 to 72 hpi, as indicated. The lower panel represents the same reactions performed with primers specific to host 28S rRNA gene.

(A)

```
                                   TATA
-500  tttcgaatca taaatcatat cgcatttata cgacacattt ttgatagatg accatgtgat ttttggcttg

-430  gtatgttgtt tgatgtattc tttcaaaaat tgaaacatgc atattggtaa cgtatcactt ggtttcaaaa

        TATA
-360  tttttaaata tattgccata ccaaaactgg cgttagcccg tgctttttt tttgttctag cgaatgcgat

          TATA
-290  tgtttctata ttttggtatt ttaacgaaca ccgaaatatt aaatctgaac ttcgtatttg ttttatgttg

                                                                  TATA
-220  acaaaatctt cttggatatc caacaacgca ccaggtaaca tgttttcatc tatatccgat aaatttattt

                       Zeste
-150  gtttgatatc aacactgagt gacatttcgg ctttgcgtga cacatattga tagaaaagtg aaacaggatc

        TATA
 -80  cgtatatatt tgaggcatac tgaaaagaac cgaacatacg ttgaacgact taacaggaca cgcgttcgtt

                 ┌▶
 -10  aatcactaaa atgccaacat gtga
```

(B)

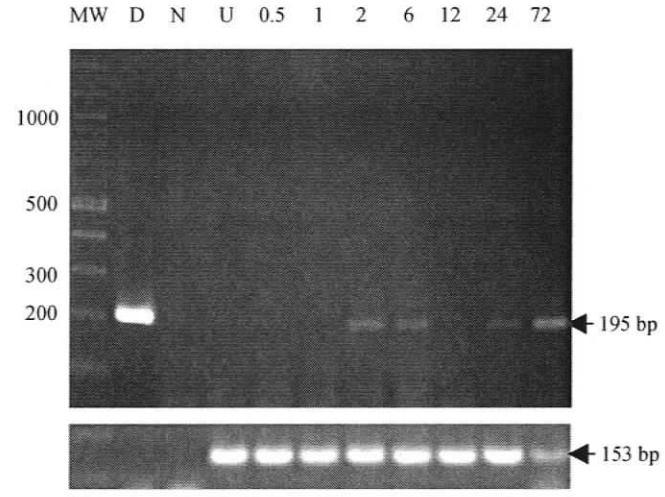


Figure 4.7. Promoter sequence and RT-PCR analysis of the *neab52* gene. (A) Promoter sequence analysis of the sequence 500 bp upstream of the predicted translation start site for the *neab52* gene. Putative TATA and Zeste transcription factor binding motifs are underlined and labelled and the predicted translation start site is in bold. (B) RT-PCR was performed from RNA extracted from uninfected and NeabNPV-infected larve. The expected size of the RT-PCR product is indicated to the right of the panel. Lane D) amplification from NeabNPV DNA (positive control) and; Lane N) amplification from no template; and Lane U) amplification from uninfected larvae (negative control). The remaining lanes are amplicons from RNA extracted from larvae 0.5 to 72 hpi, as indicated. The lower panel represents the same reactions performed with primers specific to host 28S rRNA gene.

4.5. Discussion

We have characterized two genes from the NeabNPV genome that were proposed as candidate early-expressed genes based on the identification of putative zinc finger motifs. The first gene, *neab24*, is predicted to code for a protein with a C-terminal zinc RING finger while the second gene, *neab52*, is predicted to code for a protein with four tandem N-terminal zinc finger motifs. The upstream sequences of both genes were scanned for insect host promoter elements and only TATA and Zeste elements were identified. The TATA element is common in early-expressed baculovirus genes (Guarino and Smith, 1992; Pullen and Friesen, 1995) and the Zeste motif has been proposed as an element enriched in NeabNPV early gene promoters (This thesis, Chapter 2). Transcript analysis by RT-PCR confirmed the predictions that these genes were early-expressed, with *neab24* observed within 0.5 hpi and *neab52* within 2 hpi.

Sequence analysis of the *neab24* gene product revealed a predicted RING finger domain. Although the classical zinc finger of TFIIIA (C2H2) is best characterized for its specific interaction with DNA (Engelke *et al.*, 1980; Frankel *et al.*, 1987; Hanas *et al.*, 1983), the roles of zinc fingers have since been expanded to include mediating interactions between the zinc finger and DNA, RNA and other proteins. The RING finger motif is best characterized for its function in E3 ubiquitin ligation (Lorick *et al.*, 1999). Ubiquitination has a well-established role in targeting misfolded proteins for degradation (Weissman, 2001). There is growing evidence, however, for the role of ubiquitination in the regulation of other cellular processes, including cell-cycle progression, signal transduction, transcriptional regulation, receptor down-regulation, and

endocytosis, by proteolytic degradation of regulatory proteins (reviewed in Hershko and Ciechanover, 1998).

Recent analysis of proteins in *Bombyx mori* NPV (BmMNPV) has demonstrated that IAP2, IE2 and PE38 RING finger proteins possess ubiquitin ligase E3 activity (Imai *et al.*, 2003). E3 activity was not detected for the RING finger proteins IAP1, P35, and CG30. It remains possible, however, that these proteins may have the potential for E3 activity. This is particularly true for the apoptosis inhibitors, given that their sequences are highly conserved with the mammalian X-linked IAP and c-IAP2 proteins, for which E3 activity via the RING finger has been reported (Miller, 1999; Salvesen and Duckett, 2002). These studies suggest a possible role of the neab24 RING finger proteins in ubiquitin-mediated viral pathogenesis.

There are species-specific differences between the RING finger proteins of the sawfly baculoviruses. NeabNPV and NeleNPV each have only one predicted RING finger protein, neab24 and nele8, respectively. The amino acid sequences of the neab24 and nele8 RING finger domains are conserved. NeseNPV has four RING finger proteins: nese9, nese11, nese18, and nese19. Although the nese18 and nese19 represent conserved paralogues, nese9 and nese11 do not share sequence similarity with one another, or with NeabNPV and NeleNPV RING finger proteins. This expansion of RING finger proteins in NeseNPV may represent an adaptation that facilitates host-specific pathology in *N. sertifer.*

Conversely, the *neab52* gene product encodes four C2H2 zinc finger domains. The neab52 protein shows sequence similarity to zinc fingers of the Kruppel family. Additionally, this protein possesses a HTKEKP amino acid sequence that potentially

serves as the H-C (HTGEKP) link at position 75, which is characteristic of the Kruppel family (Schuh *et al.*, 1986). Although the role of C2H2 zinc fingers in binding DNA is well established, these zinc fingers can also bind RNA or proteins (Mackay and Crossley, 1998; Shastry, 1996). Despite this, the major focus of C2H2 zinc finger research has involved protein-DNA interactions (Wolfe *et al.*, 2000). Based on these studies, it is clear that C2H2 motifs play an important role in transcription factors. In baculoviruses, where viral gene expression is primarily regulated at the level of transcription, virally encoded transcription factors play an important role in viral infection.

The IE1 protein typifies transcription regulatory proteins of the lepidopteran baculoviruses. The IE1 is a highly conserved protein and is encoded in all lepidopteran baculovirus genomes sequenced to date. An additional suspected immediate early transactivator, ME53, is also found in all lepidopteran baculoviruses. The other known immediate early transcription factors, including PE38 and IE2, are conserved only among Group II NPVs. It is unlikely that either neab24 or neab52 predicted proteins are analogous to the hallmark IE1 transactivator in lepidopteran baculoviruses. The neab24 protein, although observed within 0.5 hpi, is not conserved in NeabNPV and NeseNPV. Conversely, although neab52 is conserved among the three sawfly baculoviruses, its transcription is not observed until 2 hpi. However, both proteins may each play a significant role in regulating NeabNPV infection.

Promoter sequence elements TATA, CAKT, and DTAAG have traditionally been used in an attempt to discriminate between early and late baculovirus genes. In NeabNPV, we have previously raised questions regarding the use of CAKT as a predictor of early-expressed genes and have instead proposed that the host Zeste protein putative binding

sequence motif is enriched in early-expressed genes (This thesis, Chapter 2). Our analysis of sequences 160 bp upstream of predicted NeabNPV genes showed that the *neab52* but not the *neab24* gene was preceded by an upstream Zeste-binding sequence. This study extended this analysis to 500 bp upstream and only detected TATA and Zeste-binding insect promoter sequences. A Zeste-binding motif was observed upstream of the *neab24* gene at –345 bp. Although the minimum baculovirus promoter is typically within 120 bp of the translation start site (Dickson and Friesen, 1991; Kogan and Blissard, 1994), regulatory elements may extend upstream over hundreds of base pairs (Blissard and Rohrmann, 1991). The role of Zeste in regulating early baculovirus gene expression has not been experimentally examined, but it might help elucidate the role of host cell transcription factors in early baculovirus gene expression.

Among non-lepidopteran baculoviruses, 29 genes share significant sequence similarity to putative lepidopteran baculovirus homologues (Afonso *et al.*, 2001; Garcia-Maruniak *et al.*, 2004; Lauzon *et al.*, 2004). None of these conserved genes include the immediate early genes that are essential transcription factors are conserved elements in lepidopteran baculoviruses. Based on the well-established regulatory role of zinc finger proteins, as well as the predicted link of upstream Zeste-binding motifs to early NeabNPV genes, we predicted that the *neab24* and *neab52* genes represent novel early genes. Transcriptional analysis confirmed that these genes were transcribed early, with *neab24* transcribed within 0.5 hpi. Sequence analysis showed that, although the sequences of the gene products were not conserved with any currently sequenced baculovirus proteins, they did show zinc finger motifs with conserved characteristics that implicate them in regulation

of transcription and cellular processes. It is therefore likely that these two genes encode important NeabNPV viral regulatory proteins.

Chapter 5. Analysis of putative promoter regulatory elements of the *Baculoviridae*

5.1. Abstract

Baculovirus transcription regulation is thought to be very complex and tightly regulated. The detailed analyses of baculovirus transcriptional regulation have tended to involve only a limited subset of genes. This study addressed the need for a more global investigation of baculovirus transcription by identifying potential promoter elements by sequence alignment. The upstream sequences of baculovirus genes were remarkably conserved throughout baculovirus species. This was particularly true of early and late genes, when compared to immediate early and very late genes. Although the proximity of viral open reading frames can account for some of this sequence conservation, we identified some motifs that remain conserved even when the upstream gene does not. This suggests the possibility that baculoviral coding sequence, in addition to coding for functional proteins, may have a dual role as gene regulation elements. Support for the efficacy of upstream sequence alignment in identifying gene regulatory elements came from the identification of sequence motifs that coincide with previously reported transcription start sites. By generating Hidden Markov Models from conserved upstream motifs we showed how some sequence motifs are conserved over a wide range of different baculovirus genes and species. Some motifs also appear to by temporally biased, suggesting a possible link with temporal regulation of baculovirus genes. Although the identification of novel gene regulatory elements is a challenging task, the survey of upstream sequence motifs can allow us to take advantage of the significant

amont of available baculovirus nucleotide sequence data to identify potential novel promoter elements.

5.2. Introduction

Baculoviruses are a diverse family of viruses that have been reported to be pathogenic to over 600 species of invertebrates, mostly of the orders *Lepidoptera*, *Hymenoptera* and *Diptera*. They have relatively large genomes that encode between 89 and 181 genes (Hayakawa *et al.*, 1999; Lauzon *et al.*, 2004) and viral infection is mediated by a cascade of temporally regulated gene expression. Gene expression of the baculoviruses has been well characterized and it is clear that gene regulation is primarily regulated at the level of transcription. Although a number of key virus-encoded transcription factors (TFs) have been implicated in coordinating baculovirus transcription, there have been no global analyses comparing the significance of shared sequence elements in baculovirus promoters.

The first evidence that promoter elements significantly influenced baculovirus gene expression came from the investigation of the AcMNPV *38K* gene promoter. It was observed that cotransfection with the baculovirus gene *ie1* resulted in increased transcription from the *38K* gene promoter (Guarino and Summers, 1986). The baculovirus IE1 transactivator is expressed at a steady state throughout expression (Guarino and Summers, 1987) and is involved in regulating both early- and late-expressed baculovirus genes (Carson *et al.*, 1991; Kool *et al.*, 1994; Leisy *et al.*, 1995; Lu and Carstens, 1993; Lu and Miller, 1995; Nissen and Friesen, 1989; Passarelli and Miller, 1993; Rapp *et al.*, 1998). Other virus-encoded transactivators have been identified for the baculoviruses, including IE-2, IE-0, PE38, and ME53 (Carson *et al.*, 1991; Knebel-Morsdorf *et al.*, 1993; Kovacs *et al.*, 1991; Prikhod'ko and Miller, 1999). In contrast to

IE-1, these transregulators appear to transiently regulate gene expression during narrow temporal periods.

Another means by which baculovirus genes are regulated is by the differential specificity of host and virus-encoded RNA polymerases. Soon after the influence of the IE1 transregulator was reported, a promoter element with the consensus TAAG was identified in the highly expressed *polyhedrin* gene promoter. Deletion of this sequence resulted in 90% reduction of viral transcription from this promoter (Possee and Howard, 1987). Transcription of late and very late genes, such as polyhedrin (*polh*), initiates from the TAAG sequence motif that is specifically recognized by the α-amanitin resistant RNA polymerase encoded by four baculovirus genes, *p47*, *lef4*, *lef8*, and *lef9* (Guarino *et al.*, 1998). This temporally regulates late gene expression, as the RNA polymerase must be synthesized before late genes may be transcribed. Mutational analysis of the *polh* promoter also suggests that the seven nucleotides upstream of the TAAG sequence motif may mediate the strength of transcription from these promoters (Mans and Knebel-Morsdorf, 1998).

Transcription of the *gp64* gene has been observed both early and late in infection (Blissard and Rohrmann, 1989). Early and late transcription appear to be regulated by different promoter elements, with early expression regulated by an element at –43 bp, relative to the RNA start site, and late expression regulated by elements at -152, -167, -174, and –175 bp (Blissard and Rohrmann, 1989). Gene expression is further divided into two regions: -319 to –166 bp appears to negatively regulate transcription of *gp64*, while –166 to –77 appears to positively regulate transcription (Blissard and Rohrmann, 1991). Two TATA elements have been identified at –77 and –62 bp, but it has been

shown that the CAGT element at –43 bp, not the TATA elements, determines the site of transcription initiation (Blissard and Rohrmann, 1991; Blissard *et al.*, 1992).

In addition to virus-encoded transcription regulation, there is evidence that host cell TFs may influence baculovirus gene expression. GATA sequences in the *gp64* and *pe38* gene promoter regions are bound by insect cell host factors (Kogan and Blissard, 1994; Krappa and Knebel-Morsdorf, 1991). Additionally, an increase in the insect host transcription activator TBP has been observed following baculovirus infection, along with a nuclear redistribution of TBP during infection (Quadt *et al.*, 2002).

Although these studies have focused on fine characterization of relatively few baculovirus promoters, we have undertaken a global survey of 21 baculovirus promoters from immediate early, early, late and very late baculovirus genes.

5.3. Methods and Materials

Sequence Data. Twenty-one baculovirus genes, for which patterns of temporal transcription have been reported, were selected for this analysis. The viral orthologous clusters (VOCs) database (Hiscock and Upton, 2000) was used to identify the genomes and nucleotide location of each gene and this list was manually confirmed and edited to ensure accuracy. For each gene, upstream sequences at –1000, -500, -100, and –50 bp relative to the predicted translation start site (TSS) were obtained from GenBank as described below. Sequence alignment and sequence motif models were generated based on the genome nucleotide positions outlined in Table 5.. Histograms describing local sequence conservation in the alignment were derived by graphing the percent frequency of the most frequent nucleotide at each site in the alignment.

Baculovirus abbreviations and source sequences. The nucleopolyhedrovirus (NPV) and granulovirus (GV) upstream sequences were derived from the GenBank database: *Autographa californica* NPV (AcMNPV, NC_001623); *Rachiplusia ou* NPV (RaouNPV, NC_004323); *Bombyx mori* NPV (BmNPV, NC_001962); *Orgyia pseudotsugata* NPV (OpMNPV, NC_001875); *Choristoneura fumiferana* defective NPV (CfMNPV-def, NC_005137); *Choristoneura fumiferana* NPV (CfMNPV, NC_004778); *Epiphyas postvittana* NPV (EppoNPV, NC_003083); *Adoxophyes honmai* NPV (AdhoNPV, NC_004690); *Spodoptera litura* NPV (SpltNPV, NC_003102); *Spodoptera exigua* NPV (SpexNPV, NC_002169); *Trichoplusia ni* NPV (TnSNPV, NC_007383); *Chrysodeixis chalcites* NPV (ChchNPV, NC_007151); *Helicoverpa zea* NPV (HzNPV, NC_003349); *Mamestra configurata* NPV-A (MacoNPV-A, NC_003529); *Mamestra configurata*

NPV-B (MacoNPV-B; NC_004117), *Helicoverpa armigera* NPV (HearNPV, NC_003094); *Helicoverpa armigera* NPV-G4 (HearNPV-G4, NC_002654); *Lymantria dispar* NPV (LdMNPV, NC_001973); *Cryptophlebia leucotreta* GV (CrleGV, NC_005068); *Adoxophyes orana* GV (AdorGV, NC_005038); *Phthorimaea operculella* GV (PhopNPV, NC_004062); *Cydia pomonella* GV (CpGV, NC_002816); *Plutella xylostella* GV (PxGV, NC_002593); *Xestia c-nigrum* GV (XcGV, NC_002331); *Agrotis segetum* GV (AgseNPV, NC_005839); *Neodiprion abietis* NPV (NeabNPV, DQ317692); *Neodiprion sertifer* (NeseNPV, NC_005905); *Neodiprion lecontei* (NeleNPV, NC_005906); *Culex nigripalpus* NPV (CuniNPV, NC_003084).

ClustalW analysis of upstream sequences. As a control, 120 random sequences were generated each representing 30 sequences of lengths 50, 100, 500, and 1000 bp, respectively. An equal frequency of nucleotides was assumed. All sequences for the 21 genes were aligned using ClustalW (Thompson *et al.*, 1994), with default parameters, against their equal-length orthologues. The alignment was reported as the overall ClustalW normalized against the sequence length. The alignments were illustrated using Bioedit (Hall, 1999). The sequences were compared with 30 sequences generated randomly with the assumption of equal nucleotide frequency, with each of the lengths 50, 100, 500, and 1000 bp.

Characterization of upstream motifs. To identify potential transcription factor binding sites within a motif, the consensus sequence was queried against the TRANSFAC 6.0 public database using Patch v1.0 (Matys *et al.*, 2003). Comparison of baculovirus motifs with baculovirus upstream sequences was performed by generating a position-weighted

matrix (PWM) using HMMER v2.1.1 (Eddy, 1998). Using HMMER, a Hidden Markov Model (HMM) was generated (hmmbuild –f -nucleic) and calibrated against 5000 random sequences, assuming equal nucleotide frequency. Potential motifs were identified in each upstream sequence by comparing to the HMM (hmmsearch).

Table 5.1 Genome location of 500 bp upstream sequences used for sequence alignment and determination of sequence motifs.

Baculovirus[1]	ie0	ie1	ie2	pe38	me53	p47	helicase	dnapol	lef1	lef2	lef3
AcNPV	122832>123617	127198>128946	130857<132083	132526>133491	121205<122554	32177<33382	80694<84359	52329<55283	10513<11313	3089>3721	57721<58878
RaouNPV	120538>121323	124905>126647	128477<129700	130142>131113	118908<120260	30335<31540	78859<82524	50482<53439	9128<9931	1756>2388	55883<57040
BmNPV	112596>113381	116994>118748	120662<121930	122416>123345	110964<112319	28266<29465	72477<76145	47493<50453	5480<6292	127652>128284	52882<54039
OpNPV	117919>118656	122619>124301	128733>129950	130960>131883	116206<117573	35795<36994	77963<81634	54899<57856	9094<9825	4125<4739	60490<61611
CfNPV-def	115736>116464	120100>121797	124104<125336	127254>128210	114137<115486	30383<31567	72594<76259	49231<52212	13447>14202	1247<1858	54827<55951
CfNPV	114929>115663	119280>120962	123510<124556	125616>126650	113256<114602	31666<32865	73985<77671	49766<52738	7426<8166	1251<1859	55397<56518
EppoNPV	105706>106437	110047>111717	113541<114476	115368>116252	103917<105278	27079<28278	66775<70434	44365<47247	6427<7182	2104<2715	49791<50912
AdhoNPV	21480<22211	16102<18030			22402>23376	11592>12770	65423<69064	51193>54252	111367>112047	88017<88652	47547>49094
SpltNPV	8851>9720	14501>16567			28076<28981	34211<35479	81918<85625	63926>66994	126906<127601	110817<111449	60203>61276
SpexNPV	131937<132671	126197<128341			9261>10433	113169>114371	65980>69648	90226>93417	17829<18479	16064>16693	86788>88056
TiniSNPV	10447>10644	14983>17184			8951<10063	27903<29105	73849<77502	45127<48282	117972<118616	116005>116682	50529<51899
ChchNPV	12018>12965	16577>18805			10629<11735	30789<31988	84735<88379	51341<54487	133296<133940	131326>132006	56798<58156
HzNPV	6950>7807	11357>13324			14591>15670	29475<30713	75966<79727	59462>62524	116805<117542	111073<111789	55828>56967
MacoNPV-A	151814<152518	146746<148326			7682<8746	133218>134411	82772>86410	105862>108858	32186>32833	13841>14476	102449>103606
MacoNPV-B	155218<155922	149911<151722			7861<8925	136661>137854	85124>88753	108294>111296	30107>30754	14021>14668	104797>105978
HearNPV	6954>7811	11356>13341			14620>15699	29667<30905	75455<79216	58971>62033	116691<117428	110951<111679	55336>56475
HearNPV-G4	6949>7806	11357>13324			14603>15457	29985<30986	76050<79811	59569>62631	117333<118070	111600<112325	55934>57073
LdNPV	20538<21314	15221<16921			23670>24698	44618<45790	93899<97555	78389>81433	118724<119428	132917>133567	74856>75980
CrleGV		3303>4667		17661<18707	109943>110854	50413>51576	65371>68757	83197<86334	53711<54418	33046>33552	88685<89710
AdorGV		2754<4022			98717>99628	41567>42733	55114>58530	72618<75713	44461<45162	23834>24346	77523<78563
PhopGV		3538<4818		16369<17508	118275>119168	51966>53165	68463>71864	87378<90443	55894<56595	33154>33696	93266<94315
CpGV		3392<4858		18574<19722	122501>123412	54349>55731	70711>74106	90820<93975	58892<59599	34320>34835	96284<97345
PxGV	8955<9605	5249<6430			100218>100999	42509>43669	54064>57438	72461<75400	45219<45974	24444>25256	78816<79709
XcGV	8674<9363	4570<6024			177406>178383	74444>75628	89714>93193	121682<124978	77696<78412	29085>29654	127032<128087
AgseGV	119882>120607				3546<4475	61324>62535	52570<56040	33284>36721	66526>67287	97203<97748	30050>31078
NeabNPV						44256<45434	54804>58214	6796<9555	62660<63295	52096>52683	
NeseNPV						47747<48907	58621>62052	28065<30806	66492<67064	55903>56505	
NeleNPV						40601<41770	51083>54487	15806<18577	58925<59560	48374>48961	
CuniNPV						63234<64553	78649>82647	88220>91636	38058<38765	19839>20516	

Table 5.1. continued.

Baculovirus	lef4	lef8	lef9	vp39	p74	p6.9	odve56	gp41	vlf1	polh / granulin	p10
AcNPV	76596>77990	40523<43153	49184>50734	75534<76577	119135<121072	86712<86879	129008<130138	65607<66836	63813<64952	4520>5257	118839>119123
RaouNPV	74724>76118	38671<41301	47323>48882	73662<74705	116861<118798	84849<85016	126682<127818	63764<64993	61966<63108	3230>3967	116567>116851
BmNPV	68591>69988	35594<38227	44402>45874	67520<68572	108796<110733	80352<80549	118837<119964	59987<61198	58190<59329	1>738	108497>108709
OpNPV	73687>75060	43093<45747	51395>52864	72620<73675	112559<114493	84398<84553	124333<125457	66253<67356	64496<65620	2533<3270	
CfNPV-def	68592>69959	37187<39811	45528>47000	67549<68580	111390<113396	78756<78917	121922<122977	61298<62434	59528<60658	1<738	
CfNPV	69604>70977	38724<41348	46840>48312	68517<69593	110762<112699	81807<81959	121035<122174	61857<62942	60096<61220	1<735	
EppoNPV	62714>64093	33569<36190	41750>43222	61585<62703	101271<103205	72794<72952	111755<112873	56235<57335	54485<55606	606<1343	101008>101265
AdhoNPV	62034>63431	39209>41827	30870<32366	61097<62035	23773<25788	71167<71409	9581<10654	56568<57521	55062<56231		
SpltNPV	78360>79787	36316<39072	52839<54335	77450<78358	19706<21679	87850<88104	16693>17808	71666<72658	70133<71287		18439>18756
SpexNPV	71844<73244	108094>110814	94614<96101	73243>74223	124099<126060	63037>63264	7848>8963	79448>80443	80850>81968	1>741	123740>124006
TiniSNPV	70289>71680	29637<32474	41600>43090	69283<70290	17340>19325		7845>8921	64503<65402	62925<64103	1>741	19375<19650
ChchNPV	81183>82571	34195<37035	46131>47621	80183<81184	18970>20946		9422>10498	75409<76308	73843<75021	1>741	20991<21257
HzNPV	71988>73373	31782<34487	44670>46229	71108<71989	16195>18261	82066<82395	13377<14441	65470<66438	63812<65056	1>741	
MacoNPV-A	89049<90413	127753>130389	114085<115602	90412>91389	144165<146138	76235>76465	6400>7521	95862>96863	97207>98349	1>741	143821>144078
MacoNPV-B	91408<92772	131201>133837	116470<117963	92771>93757	147569<149542	78589>78822	6596>7717	98231>99232	99569>100711	1>741	147232>147483
HearNPV	71477>72862	31974<34679	44420>45979	70597<71478	16224>18290	81555<81884	13395<14459	64976<65944	63321<64562	1>741	
HearNPV-G4	72072>73457	32055<34760	44507>46066	71192<72073	16208>18274	82150<82479	13378<14442	65571<66539	63919<65157	1>738	
LdNPV	90212>91669	48689<51313	58726>60216	89155<90213	26645>28663	99718<100017	13685>14755	83791<84762	82303<83439	673>1410	39809>40042
CrleGV	71403<72821	101473<104070	91475>92974	72871>73722	47165<49177		11682<12743	79711>80580	80822>81949	1>747	
AdorGV	60752<62110	90017<92590	80366>81862	62173>63048	37367<39295		7045<8124	68554>69423	70672>71748	1>747	8156<8602
PhopGV	75075<76400	109360<111843	96371>97864	76455>77336	46284<48260		11464<12519	83300>84157	84637>85782	1>747	
CpGV	77033<78475	113169<115790	99306>100805	78547>79404	48578<50644		13732<14799	86083>86952	87248>88384	1>747	
PxGV	60208<61506	91271<93787	81501>82985	61542>62504	37776<39512		9646<10701	69050>69901	70196>71236	945>1691	
XcGV	101504<102847	141693<144272	132152>133633	102896>103885	71928>74060		9382<10443	113729>114601	114960>116081	1>747	
AgseGV	48527>49933	114876<117527	26745<28235	47610<48488	72262>74289		118023>119081	39556<40485	38053<39201	130934<131680	
NeabNPV	58211<59617	72378>74909	37364<38953	82641<83588	45436<47340		14061>15071	42650<43315	41078<41308	1>741	
NeseNPV	62060<63493	75562>78102	40815<42338	84855<85793	49560<51464		39165>40184	46169<47107	44850<45911	1>741	
NeleNPV	54484<55890	68351>70882	33771<35282	80133<81080	41772<43673		22321>23331	38987<39799	37649<38713	1>744	
CuniNPV	94194>95687	20567<23335	46471>48243	19004>19873	64492<66537		99162>100247	28666>29526	14847>15923	1>750	

5.4. Results

Alignment of upstream sequences. Typically, sequence alignment is used to characterize the coding sequences of genes, as they are subject to selective pressure that supports conservation of their sequences. Non-coding sequences are usually not subject to a similar degree of selective pressure and the regulatory elements of eukaryotic promoters are typically small, in the range of 10-30 bp. This can impair attempts to identify these motifs by sequence comparison.

Despite the overall sequence divergence of non-coding sequences, we hypothesized that we could discriminate between baculovirus promoters with the goal of identifying those that were more conserved overall. We limited our analysis to the identification of *cis* acting regulatory elements within 1 Kb of the TSS, assuming that the primary regulatory elements would occur upstream of the coding sequence.

We examined whether there was an overall distance at which baculovirus promoters tended to be most conserved. In general, there was no appreciable difference in overall sequence conservation based on the length of upstream sequence investigated. It did appear, however, that the 500 bp upstream sequence provided a marginally better alignment score than did shorter sequences (Figure 5.1A).

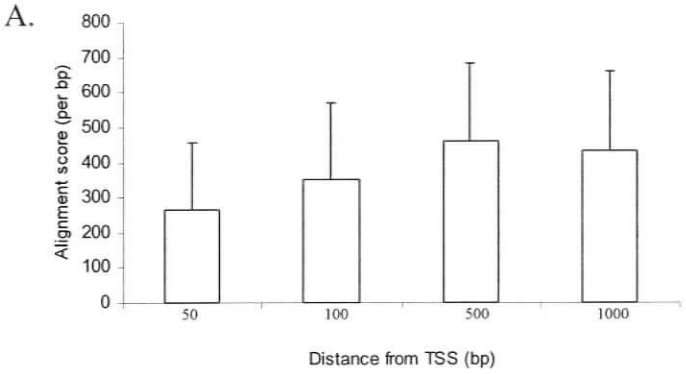
We compared the conservation between aligned sequences of distinct temporal stages. This indicated a significant increase in conservation in early- and late-expressed genes over the immediate early and very late genes (Figure 5.1B). We found that a subset of genes, *dnapol*, *lef4*, *helicase*, *vlf1* and *gp41*, that had remarkably conserved upstream sequence alignments (Figure 5.1C). Of these, *helicase* and *gp41* had an average sequence identity of 50% and 51%, respectively. These two gene upstream sequences

were not included in further analysis as the high sequence identity impaired identification of discrete conserved motifs.

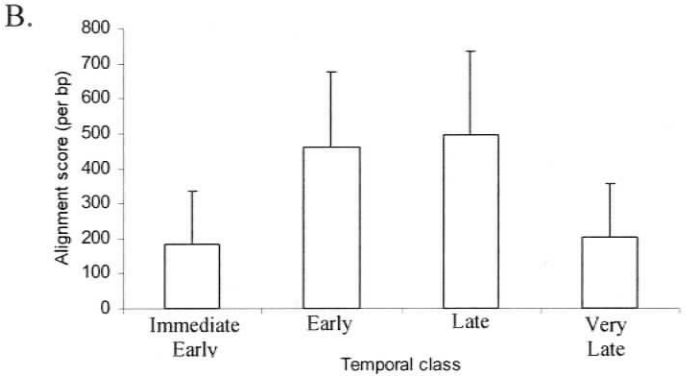
Upstream sequence of the baculovirus *DNA polymerase* (*dnapol*) gene. The regulation of the *dnapol* has been reported from a number of lepidopteran nucleopolyhedroviruses (NPVs) (Ahrens and Rohrmann, 1996; Bjornson *et al.*, 1992; Chaeychomsri *et al.*, 1995; Huang and Levin, 2001; Liu and Carstens, 1995; Tomalski *et al.*, 1988). The evidence from these studies suggests that the baculovirus *dnapol* gene utilizes a complex promoter with numerous upstream regulatory elements that coordinate its expression, and has multiple differentially regulated transcription start sites (Ahrens and Rohrmann, 1996; Bjornson *et al.*, 1992; Chaeychomsri *et al.*, 1995; Liu and Carstens, 1995; Tomalski *et al.*, 1988). Although it has been suggested that host TFs may bind upstream of the *dnapol* coding region (Huang and Levin, 2001), there is currently no experimental evidence to support this theory. These studies are also limited in that *dnapol* gene regulation has not been characterized in granuloviruses (GVs) or non-lepiopteran NPVs.

Figure 5.2A represents a histogram of the frequency of each consensus nucleotide for the aligned 500 bp upstream sequences of 29 baculovirus species. As this figure illustrates, the upstream sequence of the *dnapol* gene within 300 bp of the coding sequence is significantly more conserved that the more distal sequence. Four conserved sequence motifs were predicted by sequence alignment in the *dnapol* upstream sequences (labelled Regions I-IV).

A.


Alignment score (per bp)
800
700
600
500
400
300
200
100
0
50
100
500
1000
Distance from TSS (bp)

B.


Alignment score (per bp)
800
700
600
500
400
300
200
100
0
Immediate Early
Early
Late
Very Late
Temporal class

C.

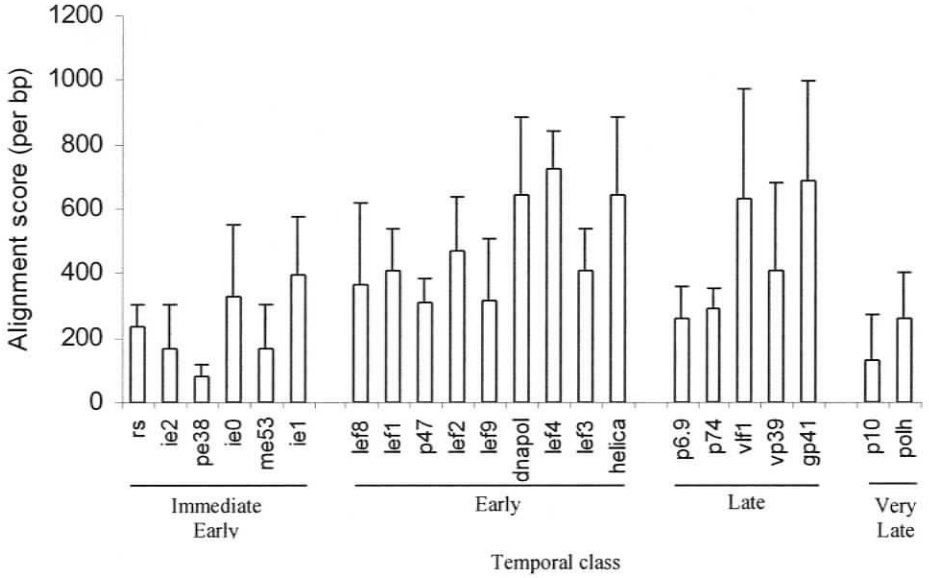
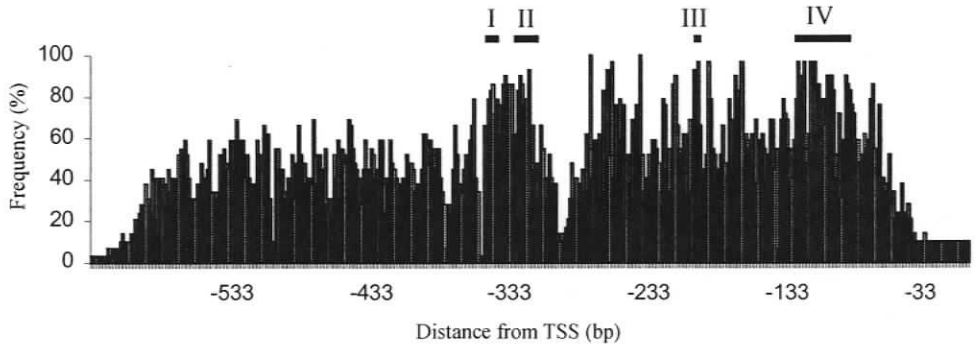

Alignment score (per bp)
1200
1000
800
600
400
200
0
rs
ie2
pe38
ie0
me53
ie1
lef8
lef1
p47
lef2
lef9
dnapol
lef4
lef3
helica
p6.9
p74
vlf1
vp39
gp41
p10
polh
Immediate Early
Early
Late
Very Late
Temporal class

Figure 5.1. Sequence similarity of aligned baculovirus upstream sequences. (A) Mean alignment score of DNA sequences upstream from all baculovirus genes in the data set at all distances 50, 100, 500, and 1000 bp from the translation start site (TSS). (B) Mean alignment score of DNA sequences upstream from all baculovirus genes in the data set, divided into temporal classes, at all distances of 50, 100, 500, and 1000 bp, from the TSS. (C) Mean alignment score, at 50, 100, 500, and 1000 bp from the TSS, for each baculovirus gene in the data set. These scores were compared with alignment of a set of random sequences (rs). The temporal class of each gene is indicated below the chart. The error bars represent the standard deviation.

Region I had a consensus sequence of TTGTAATCGT, which was conserved only within the lepidopteran NPVs and Region II had a consensus sequence of TGTA(G/A)TT(G/A)(A/T)(A/T)(A/G)TTGTG(C/T)GT. Region II is in close proximity to Region I, suggesting that the two may represent a single interrupted regulatory region. The two differ significantly, however, in that Region II is conserved in the GVs, whereas Region I is not. Neither Region I nor Region II matched potential binding sites from the TRANSFAC database.

The AcMNPV *dnapol*, and other Group I NPVs, have been described as unconventional because they lack the TATA box common in early-expressed baculovirus genes. Analysis of sequences upstream of Group II *dnapol* genes, however, revealed potential TATA Box motifs (Bjornson *et al.*, 1992; Huang and Levin, 2001). Figure 5.2B shows a conserved motif, Region III, that is a putative TATA box sequence. The conserved TATA Box motif was observation in Group II NPVs, GVs, and non-lepidopteran NPVs. The location of this TATA motif overlaps with one base of putative CAGT motifs for TnSNPV, SpexNPV, and CfMNPV. Region IV (Figure 5.2B) is a longer conserved region of the *dnapol* upstream sequences. In contrast to Regions I and II, Region IV is well conserved in both lepidopteran and non-lepidopteran baculoviruses. The Region IV sequence motif has a number of CAAT and inverted CAAT sequences at 8, 13, and 20 bp within the motif sequence.

A.



B.

Region I

TnSNPV	-273	TTGTAATCGT
AdhoNPV	-276	TTGTAATCGT
HearNPV	-338	TTGTAATCGT
HearNPV-G4	-338	TTGTAATCGT
HzNPV	-338	TTGTAATCGT
AcMNPV	-292	TTGTAATCGT
BmMNPV	-292	TTGTAATCGT
RaouNPV	-292	TTGTAATCGT
CfMNPV	-239	TTGTAATCGT
CfMNPV-def	-275	TTGTAATCAT
EppoNPV	-284	TTGTAATCGT
LdMNPV	-288	TTGTAGTCGT
OpMNPV	-239	TTGTAGTCGT
MacoNPV-A	-288	TTGTAGTCGT
MacoNPV-B	-288	TTATAATCGT
ChchNPV	-270	TTATAATCGT
SpltNPV	-267	CCGTAATCGT
SpexNPV	-282	TTATAGTCGT
XcGV	-240	TGGTTGTTTG
NeleNPV	-267	ACGTTATCAG
NeabNPV	-267	ACGTTATCGG
PhopGV	-270	CCCCATCCGT
AgseGV	-253	TTAGAACTGC
CrleGV	-264	CCCCAGGTGT
AdorGV	-261	TGCCAAGTTT
PxGV	-243	CTACTGTTAT
CpGV	-264	CCCCACTTAT
CuniNPV	-263	ACATAGTTAC
NeseNPV	-276	GTCTCGTTAC

Region II

AcMNPV	-279	TGTAGTTGAT GTTGTGCGT
BmMNPV	-279	TGTAGTTGAT GTTATGCGT
RaouNPV	-279	TGTAGTTGAT GTTGTGCGT
CfMNPV	-226	TGTAGTTTAT GTTGTGCGT
OpMNPV	-276	GGTAGTTTAT GTTGTGCGT
MacoNPV-A	-275	TGTAGTTAAA ATTGTGCGT
MacoNPV-B	-275	TGTAGTTAAA ATTGTGCGT
CfMNPV-def	-262	TGTAATTAAT GTTGTGCGT
AdhoNPV	-263	TGTAATTATA ATTATGCGT
CrleGV	-254	TATAGTTATA AGTGTGAGT
CpGV	-254	TGTAATTGTA GTTGTGTGT
HearNPV	-325	TGTAATTGTA GTTGTGTGT
HearNPV-G4	-325	TGTAATTGTA GTTGTGTGT
HzNPV	-325	TGTAATTGTA GTTGTGTGT
ChchNPV	-257	TGTAGTTGAA ATTGTGTGT
TnSNPV	-260	TGTAGTTGAA ATTATGTGT
SpexNPV	-269	TGTAGTTGAA ATTGTGTGT
LdMNPV	-275	TGTAATTGAA ATTGTGAGT
SpltNPV	-254	TGTAATTAAA ATTGTGCGT
EppoNPV	-271	TGTAGTTAAT ATTGTGTGT
XcGV	-227	CGTATTTGTA ATTGTACGT
PxGV	-230	TGTAATTGTA ATTGTAGGT
AdorGV	-251	TGTAATTGTA TGTGTGCAC
PhopGV	-260	TATAATTATA TGTGTGTGT
NeleNPV	-254	TGGCACGGAC GTCGCATTT
NeabNPV	-254	TGGCACGAAC GTCACATTT
NeseNPV	-263	TATGTCTAAT CTCGCATAT
CuniNPV	-251	CGAGCCGTAA ACTCTCCTT
AgseGV	-240	AATAGTCATG AATGCTTCC

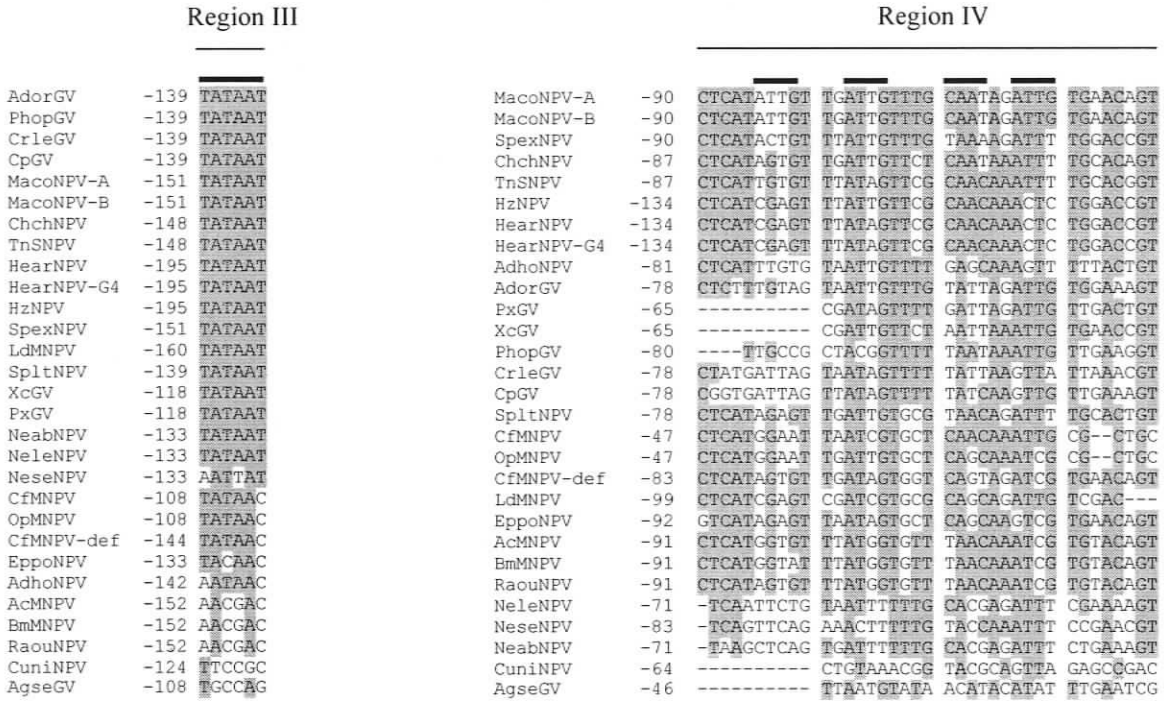


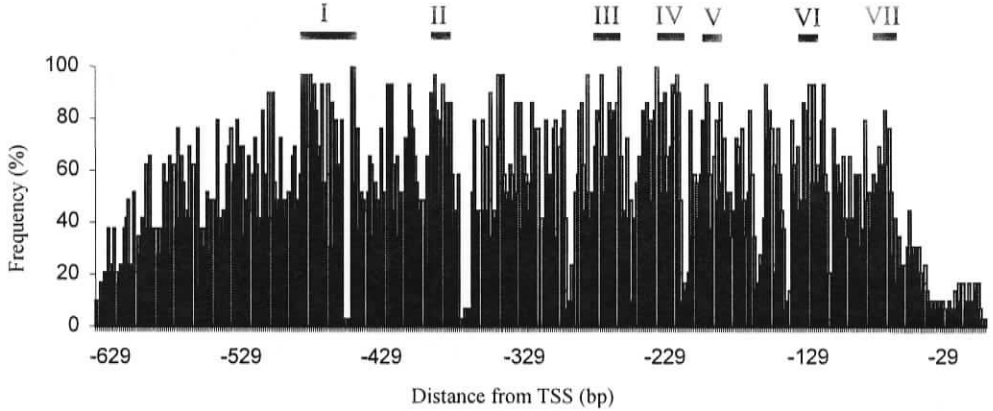
Figure 5.2. Conserved sequence motifs 500 bp upstream of the baculovirus *dnapol* gene. (A) Histogram of the percent frequency of the most frequent nucleotide, in the aligned sequences of the baculovirus *dnapol* gene, at each position relative to the transcription start site. (B) Sequence alignments of conserved nucleotide motifs in the sequence upstream of the *dnapol* gene. The position of the motif in the unaligned sequence for each baculovirus species in indicated to the left of the alignment. The motifs are labelled in roman numerals, as indicated above the alignment. The conserved sequences are highlighted in grey. The abbreviations for the baculovirus species are defined in the methods and materials. A thick horizontal bar directly above the sequence indicates a motif discussed in the text.

Upstream sequence of the baculovirus *lef-4* gene. In contrast to the *dnapol* gene, the upstream sequence of baculovirus *lef-4* genes has not been well studied. Investigations of the *lef-4* gene product have primarily focused on its role as a subunit of the equimolar, heterotetramer, baculovirus RNA polymerase (Guarino *et al.*, 1998). A significant effort has been involved in elucidating the role of LEF-4 as a capping enzyme with distinct RNA 5'-Triphosphatase and ATPase activities (Gross and Shuman, 1998; Guarino *et al.*, 1998; Jin and Guarino, 2000; Jin *et al.*, 1998; Martins and Shuman, 2001; Martins and Shuman, 2003). The interaction of this subunit complexed with a phosphate has even been resolved by crystal structure (Changela *et al.*, 2005). Despite this functional characterization, only one study has addressed transcription regulation of this gene, indicating a potential transcription start site at –56 bp (Durantel *et al.*, 1998).

As Figure 5.3 illustrates, sequence conservation in the region upstream of the *lef-4* gene is more evenly distributed than the *dnapol* upstream sequence. Within the 500 bp sequence upstream of *lef-4*, we have predicted seven conserved regions. Also in contrast to the *dnapol* upstream region, the conserved Regions I-IV of *lef-4* share a striking degree of sequence conservation, not only between the lepidopteran NPVs, but also among lepidopteran NPVs and GVs, hymenopteran NPVs, and the dipteran NPV. Conversely, Regions V and VI showed significant sequence similarity only among the lepidopteran NPVs and GV and Region VI was a short sequence conserved only among lepidopteran NPVs. Although the functional significance of the conserved Regions I-IV upstream of the *lef-4* gene is unknown, the transcription start site observed at –56 bp in AcMNPV corresponds to Region VI (Durantel *et al.*, 1998).

Sequence comparison of the conserved regions upstream of *lef-4* with TRANSFAC database indicated numerous potential binding sites similar to those of known TFs. Region I had an abdB-like Hox site (TTTACGAC) in an opposing direction to the *lef-4* ORF. The abdB-like Hox site is known to stabilize the binding of the Meis1 homeodomain protein to the motif TGACAG (Shen *et al.*, 1997). This second sequence was not, however, identified in the vicinity of the putative Hox site. Region II has sequences similar to the binding site for two zinc finger proteins, including a GATA site in the sense direction (Mackay *et al.*, 1998). The poly-A and poly-T stretches of Region III may also be targets for zinc fingers, including PF1 (TTTTT) and MIG1 (TTTA/TA) (Lundin *et al.*, 1994; Yochum and Ayer, 2001). Region IV encodes a putative GATA site at bp 2 and the consensus sequence GGNTGGT(A/T), in the antisense direction from bp 16, resembles the sequence GGGTGGTC recognized by the Zic/GLI zinc finger transcription factors (Mackay *et al.*, 1998; Mizugishi *et al.*, 2001). The conserved CGTCG sequence at bp 3 within Region V resembles the *Drosophila* Adh distal factor motif (England *et al.*, 1990). Region VI sequence resembles the CAATTATTG motif of homeodomain leucine zipper TFs (Meijer *et al.*, 2000). Region VII GGCCA sequence is similar to the target of AP-2 and NF-1 (Imagawa *et al.*, 1987).

A.



B.

Region I

CfMNPV	-394	TTGTCGTAAA TTAAATGAAA AATTAGCTGC ------TCGG
OpMNPV	-379	TTGTCGTAGA TCAAATGAAA AATCAGCTGC ------TCGG
LdMNPV	-366	TTGTCGTAGA TCATGTGCGT GACCAGCTGC ------TCGG
TnSNPV	-366	TTGTTGTAGA TGATGTGCAG AACAAGCTGT ------TCGG
CrleGV	-408	TTTTCGTAAA TAGTGTACAA GATAAACTTT ------TCCA
SpltNPV	-375	TTATTGTAAA TGAGATGCAA AACTAGGCGT ------TCTG
ChchNPV	-366	TTGTTGTAGA TGATGTGTAA CACGAGCTGT ------TCGG
HearNPV	-366	TTGTCGTAAA TCATGTAGAA CACTAACCGT ------TCGG
HearNPV-G4	-366	TTGTCGTAAA TCATGTAGAA CACTAACCGT ------TCGG
HzNPV	-366	TTGTCGTAAA TCATGTAGAA CACTAACCGT ------TCGG
AcMNPV	-389	TTGTTATATA TCAAATGAAA GATTAGCTGT ------TCGG
BmMNPV	-389	TTGTTATATA TCAAATGAAA GATTAGTTGT ------TCGG
RaouNPV	-389	TTGTTATATA TCAAATGAAA GATTAGTTGT ------TCAG
EppoNPV	-378	TTGTTATAAA TCATGTGAAT AATTAATTGT ------TCAG
SpexNPV	-366	TTATTGTAGA TCATGTGAAA CACGAGCTGT ------TCGG
CfMNPV-def	-379	TTATCATAAA TCATGTGAAA AATAAGTTGT ------TGTG
XcGV	-428	TCGTCGTAGA TCGAATAGAA CACATATTTT ------TCCA
AdorGV	-445	TCGTTGTAAA TTGTGTATAT GATCATTCTT ------TCGA
PhopGV	-434	TCTTCATATA TGGTATAAAA CACAAAACGT ------TCCA
PxGV	-391	TCCCCGTAGA TGCTGTACAT TATAAATTTT ------TCGG
CpGV	-433	TCCTCGTATA TGGTGTATAT GACAAACTTT ------TCCA
MacoNPV-A	-357	TCATCGTAAA TCATGTGGAA AATAAGCTGC ------TCCG
MacoNPV-B	-357	TGGTCGTATA TCATGTGGAA AATGAGCTGT ------TCTG
AdhoNPV	-366	TTTTCATAAA TCATGTGTAT AATCAATGCT ------TCAG
NeabNPV	-406	TTTTTCTCGA AACTTTAGAT -----ACAAA ------TCTG
NeleNPV	-403	TTTTTCTCTA AATTTCAAAT -----ACAAA ------TCTG
NeseNPV	-382	TTTTTGTAGA AAACGCATAT -----ACAAT ------TCGT
CuniNPV	-397	AGGTGGCCGT CAAGTTGAGC GGC--ACTAT ------TCGC
AgseGV	-388	TCTTTGTGGG CGCGCTTTAT TATCGATAAT ACATTTTCGG

Region II

XcGV	-307	GGTATAGTGA TACGAT
HearNPV	-282	GGTATCAAAA TACGTT
HearNPV-G4	-282	GGTATCAAAA TACGTT
HzNPV	-282	GGTATCAAAA TACGTT
SpltNPV	-291	GGTATCAGGA TGCGAT
MacoNPV-B	-273	GGTATTAAAA TGCGGT
SpexNPV	-282	GGAATCAAGA TACGTT
AcMNPV	-305	GGAATTAAAA TTCGCT
BmMNPV	-305	GGAATTAAAA TTCGCT
RaouNPV	-305	GGAATTAAAA TTCGCT
CfMNPV-def	-295	GGAATTAATA TACGCT
ChchNPV	-282	GGAATAAGTA TTCGAT
TnSNPV	-282	GGAATAAGAA TCCTTT
MacoNPV-A	-273	GGTATTAAAA CGCGGT
LdMNPV	-282	GGGATCAGGA CGCGCG
AdhoNPV	-282	GGTACCAGAA TGCGAG
CfMNPV	-310	GGAACCAGAA TGCGCT
OpMNPV	-295	GGCACCAGAA TGCGCT
EppoNPV	-294	GGCACTAAAA TTCTTT
CpGV	-355	GGTATGAGGA CTCGGT
PhopGV	-353	GGAATTAATA AACGTT
PxGV	-282	GGGATGATGA TGTTTC
CrleGV	-333	GGTATGAGAA CTCTGT
AdorGV	-367	GGGATAAGTA CTTTGT
NeleNPV	-324	TATATTGGCG TTCCGT
NeabNPV	-327	TATATTGACG TTCCAT
NeseNPV	-303	AAGATTGATA TACCAC
CuniNPV	-316	TGTACGGCGA AAGGTT
AgseGV	-298	TGTTTTGTAA ACTCTT

Region III

CrleGV	-233	TTTTTAATAT CTTGAAA
AdorGV	-255	TCTTTAAAAT TTTAAAA
SpltNPV	-188	TCTTTTCAAT TTTAAAA
XcGV	-238	CTTTTTTGAG TTTAAAA
HearNPV	-179	GTTTTTCCAT TTTAAAG
HearNPV-G4	-179	GTTTTTCCAT TTTAAAG
HzNPV	-179	GTTTTTCCAT TTTAAAG
MacoNPV-A	-176	TTTTTTCCAT TCTAAAT
MacoNPV-B	-176	TTTTTTCCAT TCTAAAT
AdhoNPV	-179	TTTTTTCCAT TTTAAAA
ChchNPV	-179	TTTTCTCCAT TTTAAAT
TnSNPV	-179	TTTTTTCCAT TTTGAAT
SpexNPV	-179	TTTTTTCCAT TTTGAAT
LdMNPV	-202	TCTTCTCGAT TTTGAAC
AcMNPV	-202	TTTTTGACAT TTTAAAA
BmMNPV	-202	TTTTTGACAT TTTAAAA
RaouNPV	-202	TTTTTGACAT TTTAAAA
CfMNPV	-207	TTTTGGCCAT TTTGAAA
OpMNPV	-192	TTTTGGACAT TTTGAAG
CfMNPV-def	-192	TTTTTGACAT TTTAAAA
EppoNPV	-191	TTTTTGACAT TTTAAAA
NeleNPV	-215	TTGTCACAGT TTTATAA
NeabNPV	-218	TGCTGATAAT TTTATAA
NeseNPV	-194	TTGCGATAGG TTTAGAA
CuniNPV	-207	TTGTAACCCA TTATAAA
PhopGV	-235	TTTTCAGTAT CCGGAAA
PxGV	-210	TTTTTTCCAA CGGAAAA
CpGV	-255	TCTTCCGTAT ACGAAAA
AgseGV	-188	AACTTAACGA AGGACAA

Region IV

```
HearNPV      -150  TGATATTCGC ACACGAACCA TCCGTC
HearNPV-G4   -150  TGATATTCGC ACACGAACCA TCCGTC
HzNPV         -90  TGATATTCGC ACACGAACCA TCCGTC
SpltNPV      -159  TGATACTGAC ACATGAACAC TCCGTC
MacoNPV-A    -147  TGATACTCGC AAATAAACCA ACCGTC
MacoNPV-B    -147  TGATACTCAC AAATGAACCA ACCGTC
AdhoNPV      -150  TGATATTCAC ATATATACCA TCCGTC
ChchNPV      -150  TGAAAATCGC AAATGAACCA ACCGTC
TnSNPV       -150  TGAAAATCGC AAATGAACCA ACCGTC
SpexNPV      -150  TGATACTCGC AAATATACCA TCCGTC
LdMNPV        -90  TGGTAGTCGC ATATGAACCA CCCGTC
AcMNPV        -90  TGGTTGTTGC AAATAAACCA TCCATC
BmMNPV       -173  TGGCTGTTGC AGATAAACCA TCCGTC
RaouNPV      -173  TGGTTGTTGC AAATAAACCA TCCATC
CfMNPV       -178  TGGTAATTGC AAATAAACCA GCCATC
OpMNPV        -90  TGGTAGTCGC AGATGAACCA GCCGTC
EppoNPV      -162  TGATAGTTAC AAATAAAGAA CCCATC
CfMNPV-def    -90  TGGTTATTGC AATAAAAATA ACCGTC
CrleGV       -204  TGATAATTAC AAACGAAAGT ACCATC
PhopGV       -206  TGAAAATTAC AAATAAACGT GCCGTC
CpGV         -226  TGGTAATTGC AAATGAAGGT GCCGTC
AdorGV       -226  TGATAGTTGC AAATAAAGGT TCCGTC
XcGV         -209  TGGTAGCTGC AAACAAACGT TCCGTC
PxGV         -181  TGATGGTTGC AGATAAAAGT GCCGTC
NeleNPV      -183  ATGTTGACAT AATCGACATA TCTGTC
NeabNPV      -186  ATGTTGACAT AATCTATATA TTTGTC
NeseNPV      -162  GAATTCACAT AATCAATATA TTTATT
CuniNPV      -175  GTATAGAAAG AAATTACAAG TCCATA
AgseGV       -156  ACATACCAAT ACAAATACAT ACCAGA
```

Region V

```
HearNPV      -114  GCGTCGGACG AACATGG
HearNPV-G4   -114  GCGTCGGACG AACATGG
HzNPV        -114  GCGTCGGACG AACATGG
ChchNPV       -58  GCATCCGGAG AACACGG
TnSNPV   -114  GCGTCCGGAG AACACGG
SpexNPV      -114  GCGTCCGGAG AGCAATT
LdMNPV        -114  GAGTCGGGCG AGCACGG
AcMNPV        -136  GCGTCGGGCG AACACGG
BmMNPV        -136  GCGTCGGGCG AACACGG
RaouNPV      -137  GCGTCGGGCG AACATGG
CfMNPV-def   -127  GCGTCGGCCG AGCACAT
OpMNPV        -127  GCGTCCGACG AGCATGG
EppoNPV      -126  GCGTCAGACG AACACGG
CpGV         -184  GCGTCCGAGC TACACTG
AdorGV       -184  GCATCTGGAC TGCAAGG
CfMNPV       -142  GCATCTGACG AGCATGG
SpltNPV      -123  GCGTCATCCG AGCACGG
MacoNPV-B    -111  GCGTCACTTG AACACGG
MacoNPV-A    -111  GCGTCATTTG AACACGG
AdhoNPV      -114  GCATCTTTGG AGCACTC
CrleGV       -173  GCATCGTTAC TACACTG
PhopGV       -164  GCGTCCTCGC TGCATTG
PxGV         -145  GCGTCTTTGC TGCAATG
XcGV         -170  TCGTCTGAAC ACGGCGG
NeseNPV      -120  AGACTTTGTA ACTCCGA
NeleNPV      -141  AAATTTTGTA ATTGTGA
NeabNPV      -144  AAATCTTGTA ATCGTGA
CuniNPV      -133  TTTTTATTCG AGGTCAA
AgseGV       -218  ATACGTCATT GACTTAT
```

Region VI

```
AdorGV       -112  AATATACAAT AATTGT---- TC
PhopGV       -104  AATATACAAT AGTTGT---- TC
XcGV         -107  AAGATACATA GATTTT---- TC
SpltNPV       -57  AATATGCAAT AGTTTT---- TC
MacoNPV-A     -51  AAAATACAAT TGTTGC---- TC
ChchNPV       -51  AAGATACAAT AATTAT---- TC
TnSNPV        -52  AAAATACAAT AATTAT---- TC
LdMNPV        -54  AAGACGCAAT AGTTTC---- TC
AcMNPV        -77  AAAATGCAGC GATTAAC-TC TC
BmMNPV        -77  AAAATGCAGC GGTTAAC-TC TC
RaouNPV       -77  AAAATGCAGC GGTTAAC-TC TC
OpMNPV        -67  AAGATGCAGT GGTTGG---- TC
CfMNPV-def    -67  ATGAGGCAGT GGTTCG---- TC
EppoNPV       -66  AAAATGCAGT GATTAG---- TC
MacoNPV-B     -51  AAAATACAGT TGTTGC---- GC
HearNPV-G4    -51  AATACGCAAA AGTTGC---- GC
HzNPV         -54  AATACGCAAA AGTTGC---- GC
HearNPV       -54  AATACGCAAA AGTTGC---- GC
PxGV          -79  AAAATGCAAT TGTTGT---- AT
AdhoNPV       -55  AAAATACAAT TACTGT---- TA
CrleGV        -99  AAAATGCAGT AATTAT---- TT
CpGV         -121  AAGACGCAGT AGTTGT---- TG
CfMNPV        -82  AAAATGCAGT GGTTGG---- TT
SpexNPV       -54  AAAATACAAT AGTTAC---- CA
CuniNPV       -66  AAAATGTATT TAAAGTCACG TT
NeleNPV       -78  TCGACACTAG TAGTGTC-TG TT
NeabNPV       -81  TCAACACCAG TAGTGTC-TG TT
NeseNPV       -57  TTCGTTGATG GTGTGTT-TG TT
AgseGV        -48  TCGGTACAAA GACAGT---- CA
```

Region VII

```
HearNPV       -8  CAAGGCCA
HearNPV-G4    -8  CAAGGCCA
HzNPV         -8  CAAGGCCA
LdMNPV        -8  CAAGTGCCA
CfMNPV       -36  CAAGCGCCA
CfMNPV-def   -21  CAAGCGCCA
SpltNPV      -11  CTAGCGCCA
ChchNPV       -8  CTAGTGCCA
TnSNPV        -8  CTAGTGCCA
AcMNPV       -28  CTAGCGCCA
RaouNPV      -28  CTAGCGCCA
OpMNPV       -21  CGAGCGCCA
EppoNPV      -20  CGAGCGCCA
AdhoNPV       -8  CGAGAGCCA
BmMNPV       -28  TTAGCGCCA
MacoNPV-A     -8  TTAGTGCCA
MacoNPV-B     -8  TTAGTGCCA
SpexNPV       -8  TCAGAGCCA
XcGV         -65  GACTGGTCA
CrleGV       -52  CA-TAGTAA
CpGV         -74  CA-TTGTAG
AdorGV       -65  CA-TTGTA-
PhopGV       -57  CA-TATTTG
PxGV         -32  AA--TGTTT
NeleNPV      -28  CACATCTAT
NeabNPV      -31  CATATCTAT
NeseNPV       -7  GA-ATATAA
CuniNPV      -15  TAGTTGTGA
AgseGV        -1  AA-------
```

Figure 5.3. Conserved sequence motifs 500 bp upstream of the baculovirus *lef-4* gene. (A) Histogram of the percent frequency of the most frequent nucleotide, in the aligned sequences of the baculovirus *lef-4* gene, at each position relative to the transcription start site. (B) Sequence alignments of conserved nucleotide motifs in the sequence upstream of the *lef-4* gene. The position of the motif in the unaligned sequence for each baculovirus species in indicated to the left of the alignment. The motifs are labelled in roman numerals, as indicated above the alignment. The conserved sequences are highlighted in grey. The abbreviations for the baculovirus species are defined in the methods and materials. A thick horizontal bar directly above the sequence indicates a motif discussed in the text.

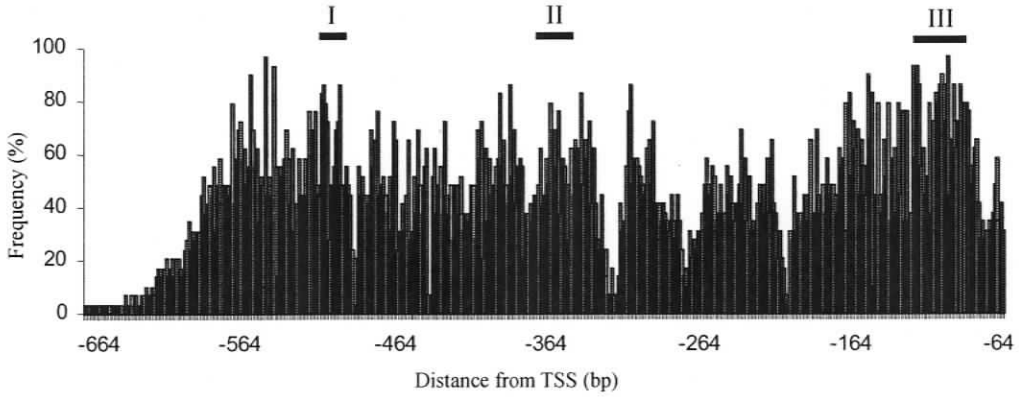
Upstream sequence of the baculovirus *vlf-1* gene. The baculovirus *vlf-1* gene is a transactivator of very late gene expression. Although transcription regulation of genes stimulated by *vlf-1* has been well characterized, little is known about the factors that stimulated *vlf-1* itself (McLachlin and Miller, 1994; Yang and Miller, 1998a; Yang and Miller, 1998b). The *vlf-1* gene transcription in AcMNPV is first observed at 12 hpi, with maximal expression from 12 to 24 hpi (McLachlin and Miller, 1994; Yang and Miller, 1998b). Transcription has been shown to initiate from a TAAG sequence at –71 bp upstream of the ORF (Yang and Miller, 1998b). Alignment of *vlf-1* upstream sequences has identified three different conserved regions (Figure 5.4).

Region I has a core TAAG sequence that is common to the start sites of late baculovirus transcripts. As at least six transcripts have been observed for the *vlf-1* gene, this motif may represent an alternate start site (Yang and Miller, 1998b). Analysis of *vlf-1* Region I with respect to TFs on TRANSFAC revealed a number of potential TF binding sites within the motif. The upstream TGTATG, at bp 1 within the motif, resembles the Actin 5C distal promoter element of *Drosophila melanogaster* (Chung and Keller, 1990). In addition, a CAGTTA sequence located at bp 6 within the motif is a binding site for the c-Myb pro-proliferation TF and a potential TATA box motif is located at bp 16 within the motif. The AATTG sequence at bp 20 within *vlf-1*.

Region II is a potential inverted CAATT site, a common binding site for transcription factors (Dorn *et al.*, 1987). Region III encodes the TAAG implicated as the site of transcription initiation in AcMNPV (Yang and Miller, 1998b), that is preceded by an AT-rich stretch that may act as a TATA motif.

Region III corresponds to the site of transcription initiation. Although the TAAG motif is conserved in the upstream sequence of lepidopteran NPV *vlf-1* genes, this sequence conservation does not extend to the GVs or non-lepidopteran NPVs. Among the upstream motif models analyzed, *vlf-1* Region III was identified in the widest range of gene upstream sequences (Table 5.1). Although the exact relationship between the motif and each upstream sequence varies, a TTTTG element is common among the upstream sequences of *dnapol*, *gp41*, PxGV *lef-9*, *p74*, AdhoNPV *polh*, and XcGV *vp39*. With the exception of NeabNPV *gp41* and PxGV *lef-9* upstream sequences, the remaining ten occurrences of TTTG associated with this motif were within 200 bp of the TSS. Contrary to the implicated role of this motif in transcription initiation, the TAAG motif was not observed in the motifs identified using the hidden Markov model.

A.



B.

Region I

SpexNPV	-426	TGCAACAGTT AAGGTTATAA A
ChchNPV	-425	TTCAACAGTT AAG-TTATAA A
TnSNPV	-437	TTCAACAGTT AAG-TTATAA A
AdhoNPV	-386	TTCAACAGTT AAG-TTATAA A
MacoNPV-A	-405	TGCAACAGTT AAG-TTACAA A
MacoNPV-B	-398	TGCAACAGTT AAG-TTACAA A
LdMNPV	-407	TTCAACAGTT AAG-TTACAA A
HearNPV-G4	-445	TTCAAAAGTT AAG-CTACAA A
HearNPV	-445	TTCAAAAGTT AAG-CTACAA A
HzNPV	-445	TTCAAAAGTT AAG-CTACAA A
SpltNPV	-455	TTCAAAAGTT AAG-CTACAA A
AcMNPV	-436	TACAATC-TT AAG--CGTAA A
RaouNPV	-437	TACAATC-TT AAG--CGTAA A
BmMNPV	-439	TACAATC-TT AAG--CGTAA A
CfMNPV	-407	TACAATC-TT AAG--CGGAA A
OpMNPV	-420	TACCATC-TT AAG--CGTAA A
EppoNPV	-422	TACAATC-TT AAG--CGTAA A
CfMNPV-def	-433	TACAATC-TT AAG--CGTAA A
AdorGV	-424	ACGAGACGTT GTAATACTAG A
PhopGV	-408	AGACGAGTTT AGACAAGTTT A
AgseGV	-425	CGCAATAATT TGACATACTC T
CpGV	-413	TGGACATAGA ACGCCCGCTC A
CrleGV	-356	TA---ACAAG ATGCCTGCTC A
XcGV	-395	TCAGCAATTG AATAGTGTGT A
PxGV	-424	GGCGGCTTTG GACAGCGTCG T
NeleNPV	-422	CGGTATCTGG AAGGAAAAAA A
NeseNPV	-409	CGTCTGCCGG AAAGAAACGA A
NeabNPV	-401	AAAGAAATAG AAAAAAACGG T
CuniNPV	-425	TCTGAAGGTG GGTTGCACGA C

Region II

AcMNPV	-289	TGGAATACAT CC--CGTTGA AATTGGCGCT
RaouNPV	-290	TGGAATACAT CC--CGTTAA AATTGGCGCT
BmMNPV	-292	TGGAATACAT CC--CGTTGA AATTGGCGCT
CfMNPV	-260	TGGATTATAT TC--CGCTCA AACTGGCGCT
OpMNPV	-273	TGGAGTACAT TC--CGCTCA AGCTGGCGCT
EppoNPV	-279	TAGAATATAT TC--CGCTCA AATTGGCACT
CfMNPV-def	-301	TCGAATACAT TC--CGCTAA AACTAGCGCT
MacoNPV-A	-265	TTCCGGA--- ----CACGAG TAATGGCGAC
MacoNPV-B	-258	TTCCGGA--- ----CACTAG CAATGGCGAA
SpexNPV	-303	TACCTGAAAA CA-ACATTAG CAGCAGCAGC
HearNPV	-299	TGGACTATAT TC--CGCTAA AATTGGCTAT
HzNPV	-299	TGGACTATAT TC--CGCTAA AATTGGCTAT
HearNPV-G4	-299	TGGACTATAT TC--CGCTAA AATTGGCTAT
LdMNPV	-285	TGATGTACAT TC--CGCTAA AATTGGCCCT
SpltNPV	-333	TCGAATACAT TC--CTTTAA AACTGGCCAT
CpGV	-265	TTGAAATTAT TGGACACTAT TAAAAACAGA
XcGV	-243	TGGACGCTAT AC--CCCTCA AATTGGCGTA
NeabNPV	-245	TCGTCCGATT CTTCCACTAA ATTCAACGAT
NeleNPV	-277	TTGAATAAGA TTT------T AATCTACAAC
NeseNPV	-264	T-GAATAAGA TTTAATGCAT AAACAACAAT
AdhoNPV	-251	TAGACGA--- ---------- AAATGGTCAT
AdorGV	-298	ATTATTTTTT TT-----TAT TTTTCCTATT
PhopGV	-262	ACTTGTCTTT ATAATAATAT ATTGTCTTTA
PxGV	-278	CTCGAAATGA AC---ACAAA AATTTATGTT
AgseGV	-273	GTCTCAGTTC CTTAGGTGCA AAATCCAAAT
ChchNPV	-280	TCGATGAAAA TGGATATAAA AAACCTAAAC
TnSNPV	-289	TTGACGAAAA TGGATACAGA AAACCCAAAC
CrleGV	-252	--AAAGTCCC TAGATA---- -AGAAACAG-
CuniNPV	-268	AGGCTCAGCG CTATTGCGAG GGCCGATGTG

Region III

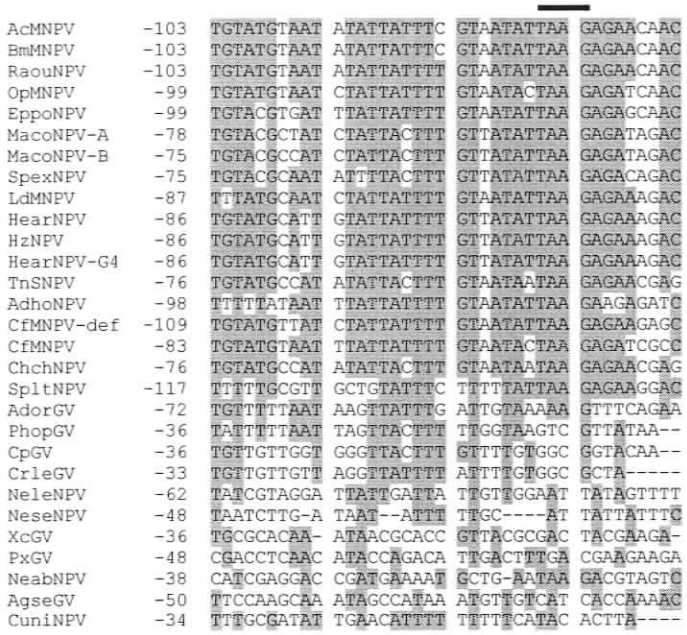


Figure 5.4. Conserved sequence motifs 500 bp upstream of the baculovirus *vlf-1* gene. (A) Histogram of the percent frequency of the most frequent nucleotide, in the aligned sequences of the baculovirus *vlf-1* gene, at each position relative to the transcription start site. (B) Sequence alignments of conserved nucleotide motifs in the sequence upstream of the *vlf-1* gene. The position of the motif in the unaligned sequence for each baculovirus species in indicated to the left of the alignment. The motifs are labelled in roman numerals, as indicated above the alignment. The conserved sequences are highlighted in grey. The abbreviations for the baculovirus species are defined in the methods and materials.

Table 5.1. Motifs identified that share local sequence similarity with the *vlf-1* Region III.

Target Gene	Species	Start	End	E-value	
dnapol	NeabNPV	-387	-360	0.067	TGAAAGTTATCCATTATTTCGGTATATT
	NeleNPV	-387	-367	4.7	TGAAAGTTATTCATTATTTCG
	SpexNPV	-41	-14	10	CGTCCGTGTTTTTATATTTTGGAACATT
	MacoNPV-A	-76	-62	8	TGTTTGCAATAGATT
	MacoNPV-B	-76	-62	8	TGTTTGCAATAGATT
gp41	AcMNPV	-31	-8	0.36	TTTTGTAATCATAAGAGTACAAAT
	BmNPV	-31	-8	0.36	TTTTGTAATCATAAGAGTACAAAT
	MacoNPV-A	-47	-26	8.6	TGTTCCTCAAATTGTATTTTGT
	MacoNPV-B	-47	-26	8.6	TGTTCCTCAAATTGTATTTTGT
	NeseNPV	-456	-417	0.034	TTTTTGTAGTGTATTATGTTTTTATCTTCATATATGTACA
	RaouNPV	-31	-8	0.36	TTTTGTAATCATAAGAGTACAAAT
	PxGV	-233	-221	6.4	TAAGAGAAAGAGC
helicase	BmNPV	-393	-367	9.2	TTATTTTTTAATGTTACAAAATGTTCA
	ChchNPV	-438	-414	2.1	TTTATGATATCAATAATTGTGTAAT
	SpltNPV	-435	-415	6.1	ATTATGCTATAATTTATTTCG
ie0	CfMNPV	-200	-184	3	TGTACGCATTATCTTAT
	HearNPV	-181	-164	1.1	TATATGCCATATTTTATT
	HearNPV-G4	-175	-158	1.1	TATATGCCATATTTTATT
	HzNPV	-171	-154	1.1	TATATGCCATATTTTATT
lef2	NeabNPV	-184	-171	8.8	TTAAGAGAAGAAAC
	AdorGV	-409	-392	5.5	TGTATGGAATTATTTATT
lef8	AcMNPV	-398	-368	1.3	TGTATTTTTTTTCAATTTTAACACATCGTCA
	BmNPV	-398	-368	1.8	TGTATTTTTTTTTAAATTTAACACATCGTCA
	RaouNPV	-398	-368	2.3	TGTATTTTTTTTCAATTTTAACACATCATCA
	AdorGV	-84	-69	5.7	TGTATACAATGTAATA
	CfMNPV	-390	-365	9	TTTTTGTAAAATATGATGCTGTCATA
	NeleNPV	-52	-32	2.4	TTTTCGCCATTTGTTAATTTG
lef9	AgseGV	-456	-437	4	AGTATGCACTTTATTATCTT
	PxGV	-392	-353	7.1	GGTTTGTTATGTGTGCAATTTTTGCATTAAGGGATGGAGC
odve56	AdhoNPV	-399	-388	5	AAGAGAAGAACT
	MacoNPV-B	-379	-364	6.5	TATTAAGAGAAGCAAA
	NeseNPV	-434	-416	2.5	TTTACACAATATATTTTTT
	PhopGV	-123	-107	2.4	TTTATACATTTTATTAT
	TnSNPV	-107	-93	3.8	TTTATGTTATATATT
p47	MacoNPV-A	-206	-186	2.2	CGTTTTCAATTCATTACTTTG
	TnSNPV	-317	-293	3.2	GATACACAATTTTTTATTTTTTTAT
	XcGV	-182	-158	9.2	TGGATGTTACGTTTTAATTCGTTTT
p74	HearNPV	-70	-31	8.5	TGTTTGTCAAAAACTTTTTTGTTCAAATAATACATGATGT
	HearNPV-G4	-70	-31	8.5	TGTTTGTCAAAAACTTTTTTGTTCAAATAATACATGATGT
	HzNPV	-70	-31	8.5	TGTTTGTCAAAAACTTTTTTGTTCAAATAATACATGATGT
	MacoNPV-A	-356	-331	8.2	TTTATGTAATCTTTTTATTTTAAATA
	SpltNPV	-350	-331	5.1	TATATGCAATTTTGTCTTTTT
polh	AdhoNPV	-125	-102	4.7	TTTTAGTAATAGTTAATTTTGTAA
	AcNPV	-345	-313	3	TTTATGTTTTTTTTATTTCATGTGATTAAGAAA
	SpexNPV	-48	-29	3.7	TGTAAGTAATTTTTTCCTTT
vp39	AdhoNPV	-147	-120	2.9	TTATTCTCATCTATTATATCATAATATT
	AdorGV	-40	-23	7.9	TGTTTGTTATTTATTGTT
	XcGV	-146	-124	0.4	TCTAGGTATTGTTTTATTTTGTA
lef-4				0.29 (Avg)	TGTATGCAAT.TATTATTTTGTAATATTAAGAGAACGACG

***DNA polymerase* Region I and *lef-4* Region IV are enriched in the upstream sequences of early-expressed genes**. The *dnapol* Region I had a conserved motif sequence identified in the early and immediate early genes, *lef-4*, *lef-8*, *lef-1*, *p47*, *ie-1*, and *me53*, of multiple baculovirus species (Table 5.2). Although the motif was also identified upstream of the late genes, *odv-e56*, *p74*, and *polh*, this observation was isolated to individual baculovirus species and did not appear to be an interspecific trend. This analysis also identified the *dnapol* motifs from which the position-weighted matrix (PWM) was generated. The average E-value obtained for these positive signals was 5.1, which is similar to the non-*dnapol* motif E-values, which ranged from 4.4-9.7. The *lef-4* Region IV motif (Table 5.3) was also observed in a number of early genes, including *lef-1*, *lef-2*, *lef-8*, and *me53*, however, it also occurred in the late genes *gp41* and *vp39*. Although enriched in the early gene upstream sequences, *lef-4* Region IV is a poor indicator of early gene promoters in comparison to *dnapol* Region I. The identified motifs also have a much higher E-value (ranging from 0.82-6.6) than the average positive signal from *the lef-4* upstream sequences (5.05E-05).

Table 5.2. Motifs that share local sequence similarity with the *dnapol* Region I.

Target Gene	Species	Start	End	E-value	
lef-4	NeabNPV	-139	-129	4.4	TTGTAATCGTG
	AdhoNPV	-210	-200	5.1	TTGTTATCGGC
	HearNPV-G4	-369	-359	7.1	TTGTTGTCGTA
	HearNPV	-369	-359	7.1	TTGTTGTCGTA
	HzNPV	-369	-359	7.1	TTGTTGTCGTA
lef-8	CrleGV	-484	-474	4.4	TTGTAATCGTG
	AdorGV	-266	-256	4.4	TTGTAATCGTT
	NeseNPV	-55	-45	4.7	TCGTAATCGTC
lef-1	NeleNPV	-438	-428	4.7	TCGTAATCGTC
	SpexNPV	-429	-419	7.7	TTGTTGTCGGT
p47	MacoNPV-B	-169	-159	4.7	TCGTAATCGTC
	SpexNPV	-263	-253	7.4	TCGTAGTCGGT
	MacoNPV-A	-187	-177	9.7	TGGTAGTCGTC
ie-1	AcNPV	-109	-99	4.8	TTGTTATCGTG
	RaouNPV	-109	-99	4.8	TTGTTATCGTG
	AdorGV	-390	-380	4.9	TTGTTATCGTT
	PxGV	-120	-110	5.4	TCGTTATCGGC
me53	MacoNPV-B	-465	-455	5.5	TCGTTATCGGA
	SpexNPV	-145	-135	7.4	TTGTTGTCGTT
odv-e56	BmNPV	-445	-435	5.1	TCGTTATCGTT
lef-2	TnSNPV	-446	-436	7	TTGTTGTCGTC
helicase	ChchNPV	-142	-132	7.2	TTGTTGTCGTG
p74	AdorGV	-205	-195	7.6	TTGTTGTCGGG
polh	AgseGV	-449	-439	4.9	TTGTTATCGTT
dnapol				5.1 (Avg)	TTGTAATCGTA

Table 5.3. Motifs identified that share local sequence similarity with the *dnapol* Region IV.

Motif Source	Motif num	Target Gene	Species	Start	End	E-value	
lef-4	IV	gp41	MacoNPV-A	-29	-3	8.20E-01	TTGTCCTCACAAACAAACACACAGTCA
			PxGV	-93	-80	7.60	TGAAAAATCCGTCC
		lef1	NeabNPV	-85	-59	5.00	TTACCGTTGCAAATAAAAGTAATATCA
			NeleNPV	-89	-63	3.90E-02	TGGTCGTTGCAAATAAAAGTAATATCA
			SpltNPV	-348	-334	2.20	TGATTTTTACAAATA
		lef2	AcNPV	-370	-355	5.60	AACAAACTAGCCATCT
			BmNPV	-370	-355	5.60	AACAAACTAACCATCT
			PhopGV	-105	-89	5.40E-01	TGATATTT---ACAAAAGAA
		lef8	CuniNPV	-177	-162	5.40	TGAGAGTGGCAAATGA
			XcGV	-249	-234	3.00	AATAAACGATGCGTCC
		me53	CfMNPV-def	-201	-189	8.50	TGCTAATTGCAAA
			XcGV	-341	-328	5.20	TGATGTTTGCAAAT
			PhopGV	-99	-83	3.90	TCATCGTTGCAAATAAA
			SpexNPV	-237	-211	1.00	CGACAATTGTATATAAAAGAACGATCT
		vp39	AgseGV	-348	-324	6.6	ATAATTTGATTTGAACAAGCCATCT
		vp39	CrleGV	-458	-435	3.3	TGGTTGCCGTTAAAGGATCCATCG
lef-4						5.05E-05 (Avg)	TGATA.TCGCAAATAAACCA.CCGTCC

5.5. Discussion

The prediction of binding sites TFs in a genome sequence is a challenging task. This challenge stems from a number of characteristics: (i) TF binding sites have small and potentially variable sequences and the degree of tolerated variability is not well understood; (ii) TF binding sites may be functional in one or both orientations and may be distal from the coding sequence; (iii) multiple TFs may coordinately or agonistically regulate the coding sequence. Some baculovirus genes, including *38K*, *gp64*, and *polh*, have received a significant degree of interest and have provided insight into the regulation and transcription of baculovirus genes (Blissard and Rohrmann, 1991; Blissard *et al.*, 1992; Kogan and Blissard, 1994; Mans and Knebel-Morsdorf, 1998; Possee and Howard, 1987; Pullen and Friesen, 1995; Rodems and Friesen, 1993). To expand our understanding of baculovirus gene regulation we compared alignments of baculovirus gene upstream sequences from 29 baculovirus genomes. We identified the five most conserved putative promoter sequences and selected three of them, *dnapol*, *lef-4*, and *vlf-1*, for more detailed analysis.

Of the three genes, *dnapol* has the best-characterized transcriptional regulation. It has been long held that the *dnapol* gene utilizes a 'TATA-less' promoter for gene transcription, based on transcriptional analysis of AcMNPV *dnapol* (Ohresser *et al.*, 1994). Our sequence alignment, however, has suggested that this may be the exception rather than the rule. Although the Group I NPVs did not appear to have a TATA element, all other baculoviruses sequenced to date have a conserved TATA motif. The TATA motif of CfMNPV overlaps with the CAGT sequence and coincides with the site of transcription initiation at –108 bp (Liu and Carstens, 1995). It is unlikely that both

sequences may serve as transcription initiation sites, however it cannot be assumed that the CAGT, and not the TATA motif, is the functional element. This contradicts transcription analysis of *Spodoptera littoralis* NPV (SpliNPV) that reports a CGT motif as the site of transcription initiation (Huang and Levin, 2001). However, baculovirus genes have long been shown to have multiple transcripts and it was speculated that SpliNPV *dnapol* may utilize multiple transcription start sites, by sequence prediction. While our observation of a TATA motif sequences may simply be coincidental, the relative significance of specific *dnapol* TATA and CAGT sequences in baculoviruses other than AcMNPV remains unclear and merits further investigation.

Our analysis indicated that early- and late- expressed gene upstream sequences were more conserved than immediate early and very late gene upstream regions. We highlighted the potential roles of conserved sequence motifs upstream of the early *dnapol* and *lef-4* genes as well as the late *vlf-1* gene. By comparing PWMs of *dnapol*, *lef-4*, and *vlf-1* conserved motifs to the upstream sequences of all genes in this analysis, we observed that the *dnapol* Region I was a common conserved motif in baculovirus upstream sequences. The *dnapol* Region I motif was also enriched in early and immediate early genes. No known TF binding sites were identified by sequence comparison of *dnapol* Region I with TRANSFAC or by literature search.

To a lesser degree, the *lef-4* Region IV motif was also enriched in early genes. The baculoviruses are a highly diverged family of viruses (Herniou *et al.*, 2004) and it would be expected that similar species would share upstream motifs, so it may be significant when motif sequences span diverged species. The *lef-4* gene Regions I, III, and IV possess motif that were unique in that they showed the highest degree of conservation

throughout the lepidopteran and non-lepidopteran baculovirus species. However, this observation is confounded somewhat by overlap between the *lef-4* upstream sequence and the upstream ORF. Overlap between upstream sequences and upstream ORFs may result in the prediction of conserved motifs that represent coding sequence motifs rather than regulatory element motifs. The coding regions of lepidopteran NPV and GV *vp39* gene begins approximately 50 bp upstream of the *lef-4* gene TSS. This association, however, is not shared with the *lef-4* ORFs of the non-lepidopteran NPVs, CuniNPV, NeleNPV, NeseNPV, and NeabNPV (Afonso *et al.*, 2001; Garcia-Maruniak *et al.*, 2004; Lauzon *et al.*, 2004; This thesis, Chapter 2). While conservation in these motifs of the non-lepidopteran NPVs supports their role as promoter regulatory elements, the potential dual role of coding sequences in some baculoviruses has not been clarified.

The *vlf-1* upstream sequence of lepidopteran baculoviruses also overlap with a common predicted ORF, ac-orf78 (Ayres *et al.*, 1994). Unlike *lef-4*, this association is observed neither in GVs nor non-lepidopteran NPVs, and this motif is only conserved throughout the lepidopteran NPVs. However, sequence comparison of this motif with the sequences 500 bp upstream of all of the 21 genes in this study revealed that this sequence motif, particularly a TTTTG sequence, is a common element in baculovirus gene upstream sequences.

This study provides an alternative approach for identifying potential gene regulatory elements in the upstream sequences of baculovirus genes. It is difficult to identify promoter elements *a priori*, based on sequence data alone, and this is particularly true for the baculovirus genomes because ORFs typically overlap with the promoters of downstream genes. However, we used sequence alignment to identify conserved motifs

in the sequences upstream of characterized baculovirus genes. Some of the predicted motifs correspond with transcription initiation of these genes and a number of the motifs may potentially bind both host and viral encoded transcription factors. Transcription analysis of baculovirus genes has typically been derived from the fine analysis of a limited subset of the genes. By coupling sequence alignment techniques with traditional transcription analyses, we can better identify the factors that mediate the complex transcription regulation of the baculoviruses.

Chapter 6. Summary and General Discussion

6.1. Revisiting Baculovirus Taxonomy and Evolution

A number of key studies have addressed the evolution and proposed taxonomy of the lepidopteran baculoviruses. These definitions have divided the *baculoviridae* into two genera, the NPVs and GVs, and these genera are further subdivided into groups based on phylogenetic evidence or viral tissue tropism, respectively. While non-lepidopteran NPVs are categorized taxanomically within the NPV genus, extensive phylogenetic analyses (Herniou *et al.*, 2004) have indicated that the NPV genus is polyphyletic and that non-lepidopteran NPVs represent distinct genera. The need for clear taxonomic classification of the *baculoviridae* family has been further spurred by the recent genome sequencing of the non-lepidopteran baculoviruses that infect the sawflies (order: *Hymenoptera*), *Neodiprion lecontei* (NeleNPV) and *Neodiprion sertifer* (NeseNPV), as well as the mosquitoes (order: *Diptera*), *Culex nigripalpus* (CuniNPV) (Afonso *et al.*, 2001; Garcia-Maruniak *et al.*, 2004; Lauzon *et al.*, 2004).

The characterization of non-lepidopteran baculoviruses is likely to expand as NPVs have been reported from the diverse orders of *Coleoptera*, *Diptera*, *Hymenoptera*, *Neuroptera*, *Trichoptera*, and possibly the *Crustacea* (Martignoni and Iwai, 1981). Molecular characterization has yet to confirm the identity of a baculovirus outside of the orders *Lepidoptera*, *Hymenoptera*, and *Diptera*. The White Spot Shrimp Virus (WSSV), isolated from the *Penaeus japonicus* shrimp (*Crustacea*), was initially classified as a baculovirus based on its morphological characteristic, but the genome sequence of this

virus is clearly unrelated to any current members of the baculovirus family (Yang *et al.*, 2001).

Chapter two of this thesis reported genome sequence for NeabNPV, the third hymenopteran baculovirus genome sequenced to date. The genome of NeabNPV was closely related to the NeleNPV and NeseNPV in genome composition, gene content, and gene order. As such, it does not directly address the taxonomy of the baculovirus family as a whole, but rather resolves the phylogeny of the sawfly baculoviruses of these related *Neodiprion* species.

Chapter two expanded on the observations from the genome sequences of NeleNPV and NeseNPV, that the hymenopteran baculoviruses represent a distinct clade from the other fully sequenced baculovirus genomes. Within the hymenopteran baculoviruses, the old world NPV (NeseNPV) was divergent from the new world NPVs (NeleNPV and NeseNPV). The divergence between baculoviruses of geographically isolated insect hosts supports the hypothesis that baculoviruses coevolve with their hosts (Herniou *et al.*, 2004). The comparison of hymenopteran baculoviruses, however, also reveals a possible common mechanism between the evolution of hymenopteran and lepidopteran baculoviruses.

A common feature of lepidopteran baculoviruses is the presence of interspersed repeats in the genome. These repeats are known as 'homologous repeat regions' (*hr*s). The *hr*s, and in particular *hr5*, have been implicated in facilitating recombination (Crouch and Passarelli, 2002). Comparison of major insertions and deletions in the *hr* loci of C1 and G4 strains of HearNPV further suggested that there was a possible association between *hr*s and the mechanisms of recombination (Zhang *et al.*, 2005). Comparison of the

NeabNPV and NeleNPV genomes of the hymenopteran baculoviruses provides the most convincing link between highly repetitive loci and genome recombination. Major rearrangements that occur between interspersed repeats in these genomes are likely to have contributed to genomic instability and facilitate the evolution of the baculoviruses. Conversely, the repeated element associated with NeabNPV and NeleNPV rearrangements was not identified in NeseNPV, suggesting that the appearance of these repeats is a relatively recent event that may have contributed to the speciation of NeabNPV and NeleNPV.

6.2. Sawfly baculovirus infection cycle

The non-lepidopteran and lepidopteran baculoviruses differ significantly in terms of pathogenesis and tissue tropism. Whereas the lepidopteran baculoviruses generally cause systemic infections, the non-lepidopteran baculoviruses are localized to the larval midgut (Federici, 1997; Moser *et al.*, 2001). While infection pathology of the sawfly NPVs is generally poorly understood, the genome sequences have aided us in comparing and contrasting lepidopteran and non-lepidopteran baculovirus pathology.

All baculoviruses sequenced to date, including NeabNPV, have a core set of 29 conserved genes (Garcia-Maruniak *et al.*, 2004; Lauzon *et al.*, 2004). These genes are involved in essential baculovirus processes, including DNA replication, transcription, and viral occlusion. Lepidopteran and sawfly baculoviruses alike have virally encoded DNA and RNA polymerase genes (Guarino *et al.*, 1998; Tomalski *et al.*, 1988). In addition to these, the baculoviruses have a host of early-expressed genes that support these mechanisms. The greatest differences in genome content, however, are observed in the

genes that are transcribed in the immediate early and in the late temporal stages of viral infection.

Immediate early baculovirus genes are best represented by the well-characterized *ie-1* gene. The *ie-1* gene is conserved in all lepidopteran baculoviruses and encodes a transactivator believed to initiate a cascade of temporally regulated gene expression that is essential for lepidopteran baculovirus DNA replication and viral propagation (Kool *et al.*, 1994). Other immediate early genes have been identified, including *ie-2*, *pe38* and *me53*, and are often essential for DNA replication, but none are as conserved throughout the baculoviridae as the *ie-1*. Although no NeabNPV immediate early genes were identified by sequence comparison with lepidopteran baculoviruses, we analyzed two NeabNPV genes, *neab24* and *neab52*, that we predicted were early-expressed. We based this prediction on two features of these genes: both genes encoded products with predicted zinc fingers, a common motif found in transcription factors and cellular regulation proteins, and both genes had an upstream Zeste protein binding motif, a feature that we found to be enriched for the early-expressed genes of the NeabNPV genome. Our analysis highlighted the potential regulatory roles of the predicted gene products, as well as confirming that they were both transcribed early in infection, with *neab24* transcribed as a potential immediate early gene at 0.5 hours post infection. Neither of these genes is likely to be analogous to the lepidopteran baculovirus *ie-1* gene as neither one is both immediate early expressed and shared among all three sawfly baculoviruses. These genes, however, are likely to each play an important role in viral infection pathogenesis.

Late viral infection in baculoviruses typically includes two distinct phenotypes: the budded virus (BV) and the occluded virus. While the occluded virus forms the occlusion bodies (OBs), or polyhedral inclusion bodies (PIBs), that are characteristic of the baculoviruses, the BV is a non-occluded virion that is responsible for cell-cell transmission of the virus (Federici, 1997) and is believed to be essential for the intertrophic invasion of the lepidopteran baculovirus. The BV is structurally similar to the nucleocapsid found in OBs but, rather than becoming occluded, it associates with and is enveloped by the host cell membrane (Volkman *et al.*, 1984). Baculovirus proteins such as GP64, or its functional analogue Ld130, are involved in cell membrane fusion and permit the cell-cell transmission of the BV (Blissard and Wenz, 1992; Monsma and Blissard, 1995). The GP64 protein is conserved among the Group I lepidopteran NPVs but was not initially identified in the basal taxon of the Group II lepidopteran NPVs, *Lymantria dispar* NPV (LdMNPV) (Kuzio *et al.*, 1999). The role of Ld130 in membrane fusion was predicted in LdMNPV based on the presence of both a N-terminal signal and transmembrane domain. An Ld130 orthologue of *Spodoptera exigua* NPV (SpexNPV), Se8, was later shown to behave as a functional analog of GP64 (IJkel *et al.*, 2000). Similarly, neither GP64 nor Ld130 were identified in non-lepidopteran baculoviruses by sequence comparison. The divergent sequences of these analogues suggest that sequence comparisons may not be the optimal means to identify potential functional homologues. As a result, we reported two predicted proteins in NeabNPV, neab10 and neab44, which putatively possess both a N-terminal signal and transmembrane domain.

Two lines of reasoning have been used to argue that the non-lepidopteran baculovirus may not have a BV phenotype, relying only on OB-transmission of the virus. The

primary support for this hypothesis is derived from the observed midgut localization of the non-lepidopteran baculovirus. It has been postulated that systemic infection, typified by the lepidopteran baculoviruses, depends on the fusion proteins that mediate BV transmission. Another line of reasoning was that the inability to develop *in vitro* propagation systems for non-lepidopteran baculoviruses stemmed from a lack of efficient cell-cell transmission via the BV phenotype. Electron microscopy has tentatively suggested, however, that the dipteran CuniNPV midgut viral transmission is mediated by a BV phenotype (Moser *et al.*, 2001). It is possible therefore, that the hymenopteran NPVs may utilize BV-mediated midgut virus transmission, similar to lepidopteran NPVs.

6.3. Impact of NeabNPV Genomics on biocontrol of the Balsam Fir Sawfly

The growing interest in sawfly baculoviruses has been spurred by the recent limited license, extended by the Canadian Pest Management Regulatory Agency (PMRA) on May 25 2005, for NeabNPV in control of the balsam fir sawfly under the trade name Abietiv™. The success of this registration package was due, in part, to our genetic characterization of NeabNPV. There has been recent extensive studies on the history of the balsam fir sawfly and the efficacy of NeabNPV aerial application (Moreau *et al.*, 2005; Moreau, 2006). Molecular analyses, however, still offer the potential to gain novel insight into the biology of NeabNPV.

The three species of *N. abietis*, *N. lectonei*, and *N. sertifer* share a common geographical origin, North America. The disappearance of the Bering land bridge isolated *N. sertifer* in Eurasia and this species became extinct in North America. Since the reintroduction of *N. sertifer* to North America, these three species once again share a

common geography. Therefore, these three insect species share a close evolutionary relationship, which is mirrored by the viruses that are pathogenic to them.

Hymenopteran baculoviruses have been reported to exhibit species-specific infection, however this has not been confirmed for NeabNPV (Federici, 1997). Differences between the sequenced genomes of the NeabNPV, NeleNPV and NeseNPV will permit the development of biomarkers that may allow detection of interspecific viral infection. Evidence of such specificity of viral pathogenicity would strongly support the registration of sawfly baculoviruses as biocontrol agents.

This study also alludes to genome elements that may be related to viral insect host or tissue trophic specificity. If the zinc finger genes are, as we have theorized, major regulators of gene transcription and/or cellular processes, they may play an important role in viral species or trophic specificity. The expansion of C2H2 ring finger proteins in NeseNPV may be a significant requirement for efficient viral propagation in *N. sertifer* midgut epithelia. The identification of individual genes that act as host range factors in baculoviruses has been previously demonstrated (Chen *et al.*, 1998) and we have observed that NeabNPV DNA replication appears to be a limiting step in attempts to propagate the virus *in vitro* (unpublished). Perhaps the lack of conservation in early genes like the zinc finger proteins may be related to host cell-specific selection pressures imposed on these genes.

NeabNPV-specific PCR primers have already been used to assay for the presence of NeabNPV in field samples (unpublished). PCR is a sensitive tool that can be used to identify NeabNPV larvae and track the epidemiology of NeabNPV in forest populations. Based on our temporal stratification of NeabNPV, RT-PCR, along with specific primer

combinations, may also be used to identify the prevalence and impact of incomplete or sublethal viral infection, which have been postulated to have a significant effect on the efficacy of bioinsecticides (Fuxa, 1987).

In general, this study provides a model for sawfly baculovirus infection that contrasts in subtle, but likely significant, ways to the model for lepidopteran baculovirus infection. It provides some possible insight into the species specificity and limited tissue trophism of the sawfly baculoviruses as well as a molecular basis to understand the divergent evolution of the sawfly and lepidopteran baculoviruses. Moreover, our molecular analyses provide powerful and senitive tools that allow us to study the sawfly baculoviruses *in vivo*.

Bibliography

A. Acharya and Gopinathan K. P. (2002). Characterization of late gene expression factors lef-9 and lef-8 from Bombyx mori nucleopolyhedrovirus. *The Journal of General Virology.* **83:** 2015-2023

C. L. Afonso, Tulman E. R., Lu Z., Balinsky C. A., Moser B. A., Becnel J. J., Rock D. L. and Kutish G. F. (2001). Genome sequence of a baculovirus pathogenic for Culex nigripalpus. *Journal of Virology.* **75:** 11157-11165

C. H. Ahrens, Carlson C. and Rohrmann G. F. (1995). Identification, sequence, and transcriptional analysis of lef-3, a gene essential for Orgyia pseudotsugata baculovirus DNA replication. *Virology.* **210:** 372-382

C. H. Ahrens and Rohrmann G. F. (1995). Replication of Orgyia pseudotsugata baculovirus DNA: lef-2 and ie-1 are essential and ie-2, p34, and Op-iap are stimulatory genes. *Virology.* **212:** 650-662

C. H. Ahrens and Rohrmann G. F. (1996). The DNA polymerase and helicase genes of a baculovirus of Orgyia pseudosugata. *The Journal of General Virology.* **77:** 825-837

S. F. Altschul, Gish W., Miller W., Myers E. W. and Lipman D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology.* **215:** 403-410

G. L. Andrews, Harris F. A., Sikorowski P. P. and McLaughlin R. E. (1973). Effects of cotton leaf surfaces on the nuclear polyhedrosis virus of Heliothis zea and Heliothis virescens. *Journal of Invertebrate Pathology.* **22:** 290-291

T. K. Attwood, Croning M. D., Flower D. R., Lewis A. P., Mabey J. E., Scordis P., Selley J. N. and Wright W. (2000). PRINTS-S: the database formerly known as PRINTS. *Nucleic Acids Research.* **28:** 225-227

M. D. Ayres, Howard S. C., Kuzio J., Lopez-Ferber M. and Possee R. D. (1994). The complete DNA sequence of Autographa californica nuclear polyhedrosis virus. *Virology.* **202:** 586-605

R. E. Balch and Bird F. E. (1944). A disease of the European spruce sawfly, *Gilpinia hercyniae* (Htg.) and its place in natural control. *Scientia Agricola.* **25:** 65-80

A. Bateman, Birney E., Durbin R., Eddy S. R., Howe K. L. and Sonnhammer E. L. (2000). The Pfam protein families database. *Nucleic Acids Research.* **28:** 263-266

A. L. Bawden, Glassberg K. J., Diggans J., Shaw R., Farmerie W. and Moyer R. W. (2000). Complete genomic sequence of the Amsacta moorei entomopoxvirus: analysis and comparison with other poxviruses. *Virology.* **274:** 120-139

D. K. Bideshi, Hice R. H., Ge B. and Federici B. A. (1998). Molecular characterization and expression of the Trichoplusia ni granulovirus helicase gene. *The Journal of General Virology.* **79:** 1309-1319

D. K. Bideshi, Renault S., Stasiak K., Federici B. A. and Bigot Y. (2003). Phylogenetic analysis and possible function of bro-like genes, a multigene family widespread among large double-stranded DNA viruses of invertebrates and bacteria. *The Journal of General Virology.* **84:** 2531-2544

F. T. Bird and Elgee D. E. (1957). A virus disease and introduced parasites as factors controlling the European spruce sawfly. *Canadian Entomologist.* **89:** 371-378

F. T. Bird and Whalen M. M. (1953). A virus disease of the European Pine Sawfly, Neodiprion sertifer (Geoffr.). *Canandian Entomologist.* **85:** 437-446

R. M. Bjornson, Glocker B. and Rohrmann G. F. (1992). Characterization of the nucleotide sequence of the Lymantria dispar nuclear polyhedrosis virus DNA polymerase gene region. *The Journal of General Virology.* **73 (Pt 12):** 3177-3183

G. W. Blissard. (1996). Baculovirus--insect cell interactions. *Cytotechnology.* **20:** 73-93

G. W. Blissard, Kogan P. H., Wei R. and Rohrmann G. F. (1992). A synthetic early promoter from a baculovirus: roles of the TATA box and conserved start site CAGT sequence in basal levels of transcription. *Virology.* **190:** 783-793

G. W. Blissard, Quant-Russell R. L., Rohrmann G. F. and Beaudreau G. S. (1989). Nucleotide sequence, transcriptional mapping, and temporal expression of the gene encoding p39, a major structural protein of the multicapsid nuclear polyhedrosis virus of

Orgyia pseudotsugata. *Virology.* **168:** 354-362

G. W. Blissard and Rohrmann G. F. (1989). Location, sequence, transcriptional mapping, and temporal expression of the gp64 envelope glycoprotein gene of the Orgyia pseudotsugata multicapsid nuclear polyhedrosis virus. *Virology.* **170:** 537-555

G. W. Blissard and Rohrmann G. F. (1990). Baculovirus diversity and molecular biology. *Annual Review of Entomology.* **35:** 127-155

G. W. Blissard and Rohrmann G. F. (1991). Baculovirus gp64 gene expression: analysis of sequences modulating early transcription and transactivation by IE1. *Journal of Virology.* **65:** 5820-5827

G. W. Blissard and Wenz J. R. (1992). Baculovirus gp64 envelope glycoprotein is sufficient to mediate pH-dependent membrane fusion. *Journal of Virology.* **66:** 6829-6835

K. L. Borden and Freemont P. S. (1996). The RING finger domain: a recent example of a sequence-structure family. *Current Opinion in Structural Biology.* **6:** 395-401

S. C. Braunagel, Elton D. M., Ma H. and Summers M. D. (1996a). Identification and analysis of an Autographa californica nuclear polyhedrosis virus structural protein of the occlusion-derived virus envelope: ODV-E56. *Virology.* **217:** 97-110

S. C. Braunagel, He H., Ramamurthy P. and Summers M. D. (1996b). Transcription, translation, and cellular localization of three Autographa californica nuclear polyhedrosis virus structural proteins: ODV-E18, ODV-E35, and ODV-EC27. *Virology.* **222:** 100-114

S. C. Braunagel, Parr R., Belyavskyi M. and Summers M. D. (1998). Autographa californica nucleopolyhedrovirus infection results in Sf9 cell cycle arrest at G2/M phase. *Virology.* **244:** 195-211

D. W. Buchan, Shepherd A. J., Lee D., Pearl F. M., Rison S. C., Thornton J. M. and Orengo C. A. (2002). Gene3D: structural assignment for whole genes and genomes using the CATH domain structure database. *Genome Research.* **12:** 503-514

M. P. Carpes, de Castro M. E., Soares E. F., Villela A. G., Pinedo F. J. and Ribeiro B. M. (2005). The inhibitor of apoptosis gene (iap-3) of Anticarsia gemmatalis multicapsid

nucleopolyhedrovirus (AgMNPV) encodes a functional IAP. *Archives of Virology.* **150:** 1549-1562

W. J. Carroll. (1962). Some Aspects of the *Neodiprion abietis* (Harr.) complex in Newfoundland. *Ph. D thesis. State University College of Forestry, Syracuse University.*

D. D. Carson, Guarino L. A. and Summers M. D. (1988). Functional mapping of an AcNPV immediately early gene which augments expression of the IE-1 trans-activated 39K gene. *Virology.* **162:** 444-451

D. D. Carson, Summers M. D. and Guarino L. A. (1991). Transient expression of the Autographa californica nuclear polyhedrosis virus immediate-early gene, IE-N, is regulated by three viral elements. *Journal of Virology.* **65:** 945-951

E. B. Carstens, Liu J. J. and Dominy C. (2002). Identification and molecular characterization of the baculovirus CfMNPV early genes: ie-1, ie-2 and pe38. *Virus Research.* **83:** 13-30

E. B. Carstens, Lu A. L. and Chan H. L. (1993). Sequence, transcriptional mapping, and overexpression of p47, a baculovirus gene regulating late gene expression. *Journal of Virology.* **67:** 2513-2520

S. Chaeychomsri, Ikeda M. and Kobayashi M. (1995). Nucleotide sequence and transcriptional analysis of the DNA polymerase gene of Bombyx mori nuclear polyhedrosis virus. *Virology.* **206:** 435-447

A. Changela, Martins A., Shuman S. and Mondragon A. (2005). Crystal structure of baculovirus RNA triphosphatase complexed with phosphate. *Journal of Biological Chemistry.* **280:** 17848-17856

C.A. Charlton and Volkman, L.E. (1993). Penetration of Autographa californica nuclear polyhedrosis virus nucleocapsids into IPLB Sf 21 cells induces actin cable formation. *Virology.* **197:** 245-254.

C. J. Chen, Quentin M. E., Brennan L. A., Kukel C. and Thiem S. M. (1998). Lymantria dispar nucleopolyhedrovirus hrf-1 expands the larval host range of Autographa californica nucleopolyhedrovirus. *Journal of Virology.* **72:** 2526-2531

T. Chen, Sahri D. and Carstens E. B. (2004). Characterization of the interaction between P143 and LEF-3 from two different baculovirus species: Choristoneura fumiferana nucleopolyhedrovirus LEF-3 can complement Autographa californica nucleopolyhedrovirus LEF-3 in supporting DNA replication. *Journal of Virology.* **78:** 329-339

G. E. Chisholm and Henner D. J. (1988). Multiple early transcripts and splicing of the Autographa californica nuclear polyhedrosis virus IE-1 gene. *Journal of Virology.* **62:** 3193-3200

J. Y. Choi, Kim Y. L. and Yang J. M. (1998). Molecular cloning and analysis of transcription initiation in the Anagrapha falcifera multiple nucleocapsid polyhedrosis virus polyhedrin gene. *Molecules and Cells.* **8:** 537-543

C. M. Chou, Huang C. J., Lo C. F., Kou G. H. and Wang C. H. (1996). Characterization of Perina nuda nucleopolyhedrovirus (PenuNPV) polyhedrin gene. *Journal of Invertebrate Pathology.* **67:** 259-266

Y. T. Chung and Keller E. B. (1990). Positive and negative regulatory elements mediating transcription from the Drosophila melanogaster actin 5C distal promoter. *Molecular and Cellular Biology.* **10:** 6172-6180

J. P. Condreay, Witherspoon S. M., Clay W. C. and Kost T. A. (1999). Transient and stable gene expression in mammalian cells transduced with a recombinant baculovirus vector. *Proceedings of the National Academy of Sciences of the United States of America.* **96:** 127-132

L. G. Copping and Menn J. J. (2000). Biopesticides: a review of their action, applications and efficacy. *Pest Management Science.* **56:** 651-676

F. Corpet, Gouzy J. and Kahn D. (1999). Recent improvements of the ProDom database of protein domain families. *Nucleic Acids Research.* **27:** 263-267

J. S. Cory, Hirst M. L., Williams T., Halls R. S., Goulson D., Green B. M., Carty T. M., Posee R. D., Cayley P. J. and Bishop D. H. L. (1994). Field trial of a genetically improved baculovirus insecticide. *Nature.* **370:** 138-140

E. A. Crouch and Passarelli A. L. (2002). Genetic requirements for homologous recombination in Autographa californica nucleopolyhedrovirus. *Journal of Virology.* **76:**

9323-9334

W. A. L. David, Ellaby S. J. and Gardiner B. O. C. (1971). The stabilizing effect of insect hemolymph on a granulosis virus held in darkness as dry films in intact capsules. *Journal of Invertebrate Pathology.* **17:** 404-409

D. De Cesare and Sassone-Corsi P. (2000). Transcriptional regulation by cyclic AMP-responsive factors. *Progress in Nucleic Acid Research and Molecular Biology.* **64:** 343-369

J. A. Dickson and Friesen P. D. (1991). Identification of upstream promoter elements mediating early transcription from the 35,000-molecular-weight protein gene of Autographa californica nuclear polyhedrosis virus. *Journal of Virology.* **65:** 4006-4016

A. Dorn, Bollekens J., Staub A., Benoist C. and Mathis D. (1987). A multiplicity of CCAAT box-binding proteins. *Cell.* **50:** 863-872

I. W. Duncan. (2002). Transvection effects in Drosophila. *Annual Review of Genetics.* **36:** 521-556

D. Durantel, Croizier G., Ravallec M. and Lopez-Ferber M. (1998). Temporal expression of the AcMNPV lef-4 gene and subcellular localization of the protein. *Virology.* **241:** 276-284

S. R. Eddy. (1998). Profile hidden Markov models. *Bioinformatics (Oxford, England).* **14:** 755-763

D. R. Engelke, Ng S. Y., Shastry B. S. and Roeder R. G. (1980). Specific interaction of a purified transcription factor with an internal control region of 5S RNA genes. *Cell.* **19:** 717-728

B. P. England, Heberlein U. and Tjian R. (1990). Purified Drosophila transcription factor, Adh distal factor-1 (Adf-1), binds to sites in several Drosophila promoters and activates transcription. *Journal of Biological Chemistry.* **265:** 5086-5094

J. T. Evans, Leisy D. J. and Rohrmann G. F. (1997). Characterization of the interaction between the baculovirus replication factors LEF-1 and LEF-2. *Journal of Virology.* **71:**

3114-3119

B. Ewing and Green P. (1998). Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Research.* **8:** 186-194

B. Ewing, Hillier L., Wendl M. C. and Green P. (1998). Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Research.* **8:** 175-185

O. Faktor, Toister-Achituv M., Nahum O. and Kamensky B. (1997). The p10 gene of Spodoptera littoralis nucleopolyhedrovirus: nucleotide sequence, transcriptional analysis and unique gene organization in the p10 locus. *The Journal of General Virology.* **78:** 2119-2128

B. A. Federici. (1993). Viral pathobiology in relation to insect control (N. E. Beckage, S. N. Thompson and B. A. Federici.) in Pathogens. Academic Press, San Diego.

B. A. Federici. (1997). Baculovirus Pathogenesis (L. K. Miller.) in The Baculoviruses. 1st ed., Plenum Press, New York.

B. A. Federici and Stern V. M. (1990). Replication and occlusion of a granulosis virus in larval and adult midgut epithelium of the western grapeleaf skeletonizer, Harrisina brillians. *Journal for Invertebrate Pathology.* **56:** 401-414

J. T. M. Flipsen, Martens J. W. M., Van Oers M. M., Vlak J. M. and Van Lent,J. W. M. (1995). Passage of *Autographa californica* Nuclear Polyhedrosis Virus through the Midgut Epithelium of *Spodoptera exigua* Larvae. *Virology.* **208:** 328-335

A. D. Frankel, Berg J. M. and Pabo C. O. (1987). Metal-dependent folding of a single zinc finger from transcription factor IIIA. *Proceedings of the National Academy of Sciences of the United States of America.* **84:** 4841-4845

L. Y. Fuchs, Woods M. S. and Weaver R. F. (1983). Viral Transcription During Autographa californica Nuclear Polyhedrosis Virus Infection: a Novel RNA Polymerase Induced in Infected Spodoptera frugiperda Cells. *Journal of Virology.* **48:** 641-646

J. R. Fuxa. (1987). Ecological Considerations for the use of Entomopathogens in Ipm. *Annual Review of Entomology.* **32:** 225-251

A. Garcia-Maruniak, Maruniak J. E., Zanotto P. M., Doumbouya A. E., Liu J. C., Merritt T. M. and Lanoie J. S. (2004). Sequence analysis of the genome of the Neodiprion sertifer nucleopolyhedrovirus. *Journal of Virology.* **78:** 7036-7051

S. Garcia-Vallve, Romeu A. and Palau J. (2000a). Horizontal gene transfer in bacterial and archaeal complete genomes. *Genome Research.* **10:** 1719-1725

S. Garcia-Vallve, Romeu A. and Palau J. (2000b). Horizontal gene transfer of glycosyl hydrolases of the rumen fungi. *Molecular Biology and Evolution.* **17:** 352-361

R. W. Glaser and Chapman J. W. (1913). The wilt disease of gypsy moth caterpillars. *Journal of Economic Entomology.* **6:** 479-488

B. Glocker, Hoopes R. R.,Jr and Rohrmann G. F. (1992). In vitro transactivation of baculovirus early genes by nuclear extracts from Autographa californica nuclear polyhedrosis virus-infected Spodoptera frugiperda cells. *Journal of Virology.* **66:** 3476-3484

J. Gough, Karplus K., Hughey R. and Chothia C. (2001). Assignment of homology to genome sequences using a library of hidden Markov models that represent all proteins of known structure. *Journal of Molecular Biology.* **313:** 903-919

C. H. Gross and Shuman S. (1998). RNA 5'-triphosphatase, nucleoside triphosphatase, and guanylyltransferase activities of baculovirus LEF-4 protein. *Journal of Virology.* **72:** 10020-10028

M. A. Grula, Buller P. L. and Weaver R. F. (1981). α-Amanitin-Resistant Viral RNA Synthesis in Nuclei Isolated from Nuclear Polyhedrosis Virus-Infected *Heliothis zea* Larvae and *Spodoptera frugiperda* Cells. *Journal of Virology.* **38:** 916-921

L. A. Guarino and Summers M. D. (1987). Nucleotide sequence and temporal expression of a baculovirus regulatory gene. *Journal of Virology.* **61:** 2091-2099

L. A. Guarino and Summers M. D. (1986). Interspersed Homologous DNA of Autographa californica Nuclear Polyhedrosis Virus Enhances Delayed-Early Gene Expression. *Journal of Virology.* **60:** 215-223

L. A. Guarino, Gonzalez M. A. and Summers M. D. (1986). Complete Sequence and Enhancer Function of the Homologous DNA Regions of Autographa californica Nuclear Polyhedrosis Virus. *Journal of Virology.* **60:** 224-229

L. A. Guarino, Jin J. and Dong W. (1998). Guanylyltransferase activity of the LEF-4 subunit of baculovirus RNA polymerase. *Journal of Virology.* **72:** 10003-10010

L. A. Guarino and Smith M. (1992). Regulation of delayed-early gene transcription by dual TATA boxes. *Journal of Virology.* **66:** 3733-3739

L. A. Guarino and Summers M. D. (1986). Functional mapping of a trans-activating gene required for expression of a baculovirus delayed-early gene. *Journal of Virology.* **57:** 563-571

L. A. Guarino, Xu B., Jin J. and Dong W. (1998). A virus-encoded RNA polymerase purified from baculovirus-infected cells. *Journal of Virology.* **72:** 7985-7991

S. Gutierrez, Kikhno I. and Lopez Ferber M. (2004). Transcription and promoter analysis of pif, an essential but low-expressed baculovirus gene. *The Journal of General Virology.* **85:** 331-341

D. H. Haft, Loftus B. J., Richardson D. L., Yang F., Eisen J. A., Paulsen I. T. and White O. (2001). TIGRFAMs: a protein family resource for the functional identification of proteins. *Nucleic Acids Research.* **29:** 41-43

T. Hai and Hartman M. G. (2001). The molecular biology and nomenclature of the activating transcription factor/cAMP responsive element binding family of transcription factors: activating transcription factor proteins and homeostasis. *Gene.* **273:** 1-11

T. Hai, Wolfgang C. D., Marsee D. K., Allen A. E. and Sivaprasad U. (1999). ATF3 and stress responses. *Gene Expression.* **7:** 321-335

T. W. Hai, Liu F., Coukos W. J. and Green M. R. (1989). Transcription factor ATF cDNA clones: an extensive family of leucine zipper proteins able to selectively form DNA-binding heterodimers. *Genes & Development.* **3:** 2083-2090

T. A. Hall. (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series.* **41:** 95-98

J. S. Hanas, Hazuda D. J., Bogenhagen D. F., Wu F. Y. and Wu C. W. (1983). Xenopus transcription factor A requires zinc for binding to the 5 S RNA gene. *Journal of Biological Chemistry.* **258:** 14120-14125

Y. Hashimoto, Hayakawa T., Ueno Y., Fujita T., Sano Y. and Matsumoto T. (2000). Sequence analysis of the Plutella xylostella granulovirus genome. *Virology.* **275:** 358-372

T. Hayakawa, Ko R., Okano K., Seong S. I., Goto C. and Maeda S. (1999). Sequence analysis of the Xestia c-nigrum granulovirus genome. *Virology.* **262:** 277-297

J. G. Heldens, Liu Y., Zuidema D., Goldbach R. W. and Vlak J. M. (1998). A highly conserved genomic region in baculoviruses: sequence analysis of an 11.3 kbp DNA fragment (46.5-55.1 m.u.) of the Spodoptera exigua multicapsid nucleopolyhedrovirus. *Virus research.* **55:** 187-198

E. A. Herniou, Luque T., Chen X., Vlak J. M., Winstanley D., Cory J. S. and O'Reilly D. R. (2001). Use of whole genome sequence data to infer baculovirus phylogeny. *Journal of Virology.* **75:** 8117-8126

E. A. Herniou, Olszewski J. A., Cory J. S. and O'Reilly D. R. (2003). The genome sequence and evolution of baculoviruses. *Annual Review of Entomology.* **48:** 211-234

E. A. Herniou, Olszewski J. A., O'Reilly D. R. and Cory J. S. (2004). Ancient coevolution of baculoviruses and their insect hosts. *Journal of Virology.* **78:** 3244-3251

A. Hershko and Ciechanover A. (1998). The ubiquitin system. *Annual Review of Biochemistry.* **67:** 425-479

D. Hiscock and Upton C. (2000). Viral Genome DataBase: storing and analyzing genes and proteins from complete viral genomes. *Bioinformatics (Oxford, England).* **16:** 484-485

K. Hofmann, Bucher P., Falquet L. and Bairoch A. (1999). The PROSITE database, its status in 1999. *Nucleic Acids Research.* **27:** 215-219

T. Hong, Braunagel S. C. and Summers M. D. (1994). Transcription, translation, and cellular localization of PDV-E66: a structural protein of the PDV envelope of Autographa californica nuclear polyhedrosis virus. *Virology.* **204:** 210-222

R. R. Hoopes and Rohrmann G. F. (1991). In vitro transcription of baculovirus immediate early genes: accurate mRNA initiation by nuclear extracts from both insect and human cells. *Proceedings of the National Academy of Sciences of the United States of America.* **88:** 4513-4517

K. Hoover, Schultz C. M., Lane S. S., Bonning B. C., Duffey S. S., McCutchen B. F. and Hammock B. D. (1995). Reduction in damage to cotton plants by a recombinant baculovirus that knocks moribound larvae of Heliothis virescens off the plant. *Biological Control.* **5:** 419-426

J. Huang and Levin D. B. (2001). Identification, transcription and sequence analysis of the Spodoptera littoralis nucleopolyhedrovirus (SpliNPV) DNA polymerase gene. *Archives of Virology.* **146:** 303-326

J. D. Hughes, Estep P. W., Tavazoie S. and Church G. M. (2000). Computational identification of cis-regulatory elements associated with groups of functionally related genes in Saccharomyces cerevisiae. *Journal of Molecular Biology.* **296:** 1205-1214

N. E. Huh and Weaver R. F. (1990). Identifying the RNA polymerases that synthesize specific transcripts of the Autographa californica nuclear polyhedrosis virus. *The Journal of General Virology.* **71:** 195-201

O. Hyink, 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. (2002). Whole genome analysis of the Epiphyas postvittana nucleopolyhedrovirus. *The Journal of General Virology.* **83:** 957-971

W. F. IJkel, van Strien E. A., Heldens J. G., Broer R., Zuidema D., Goldbach R. W. and Vlak J. M. (1999). Sequence and organization of the Spodoptera exigua multicapsid nucleopolyhedrovirus genome. *The Journal of General Virology.* **80:** 3289-3304

W. F. IJkel, Westenberg M., Goldbach R. W., Blissard G. W., Vlak J. M. and Zuidema D. (2000). A novel baculovirus envelope fusion protein with a proprotein convertase cleavage site. *Virology.* **275:** 30-41

M. Ikeda, Yanagimoto K. and Kobayashi M. (2004). Identification and functional analysis of Hyphantria cunea nucleopolyhedrovirus iap genes. *Virology.* **321:** 359-371

M. Imagawa, Chiu R. and Karin M. (1987). Transcription factor AP-2 mediates induction by two different signal-transduction pathways: protein kinase C and cAMP. *Cell.* **51:** 251-260

N. Imai, Matsuda N., Tanaka K., Nakano A., Matsumoto S. and Kang W. (2003). Ubiquitin ligase activities of Bombyx mori nucleopolyhedrovirus RING finger proteins. *Journal of Virology.* **77:** 923-930

N. J. Jakob, Muller K., Bahr U. and Darai G. (2001). Analysis of the first complete DNA sequence of an invertebrate iridovirus: coding strategy of the genome of Chilo iridescent virus. *Virology.* **286:** 182-196

R. P. Jaques. (1971). Tests on protectants for foliar deposits of a polyhedrosis virus. *Journal of Invertebrate Pathology.* **17:** 9-16

R. P. Jaques. (1972). The inactivation of foliar deposits of viruses of Trichoplusia ni and Pieris rapae and test on protectant activities. *Journal of Invertebrate Pathology.* **104:** 1985-1994

D. L. Jarvis. (1997). Baculovirus expression vectors (L. K. Miller.) in The Baculoviruses. Plenum Press, New York.

J. Jin, Dong W. and Guarino L. A. (1998). The LEF-4 subunit of baculovirus RNA polymerase has RNA 5'-triphosphatase and ATPase activities. *Journal of Virology.* **72:** 10011-10019

J. Jin and Guarino L. A. (2000). 3'-end formation of baculovirus late RNAs. *Journal of Virology.* **74:** 8930-8937

J. A. Kennison and Southworth J. W. (2002). Transvection in Drosophila. *Advances in Genetics.* **46:** 399-420

D. Knebel-Morsdorf, Kremer A. and Jahnel F. (1993). Baculovirus gene ME53, which contains a putative zinc finger motif, is one of the major early-transcribed genes. *Journal*

of Virology. **67:** 753-758

D. L. Knudson and Tinsley T. W. (1978). Replication of a nuclear polyhedrosis virus in a continuous cell line of Spodoptera frugiperda: partial characterization of the viral DNA, comparative DNA-DNA hybridization, and patterns of DNA synthesis. *Virology.* **87:** 42-57

P. H. Kogan and Blissard G. W. (1994). A baculovirus gp64 early promoter is activated by host transcription factor binding to CACGTG and GATA elements. *Journal of Virology.* **68:** 813-822

M. Kool, Ahrens C. H., Goldbach R. W., Rohrmann G. F. and Vlak J. M. (1994). Identification of genes involved in DNA replication of the Autographa californica baculovirus. *Proceedings of the National Academy of Sciences of the United States of America.* **91:** 11212-11216

M. Kool, Voeten J. T., Goldbach R. W., Tramper J. and Vlak J. M. (1993). Identification of seven putative origins of Autographa californica multiple nucleocapsid nuclear polyhedrosis virus DNA replication. *The Journal of General Virology.* **74:** 2661-2668

G. R. Kovacs, Choi J., Guarino L. A. and Summers M. D. (1992). Functional dissection of the Autographa californica nuclear polyhedrosis virus immediate-early 1 transcriptional regulatory protein. *Journal of Virology.* **66:** 7429-7437

G. R. Kovacs, Guarino L. A. and Summers M. D. (1991). Novel regulatory properties of the IE1 and IE0 transactivators encoded by the baculovirus Autographa californica multicapsid nuclear polyhedrosis virus. *Journal of Virology.* **65:** 5281-5288

R. Krappa and Knebel-Morsdorf D. (1991). Identification of the very early transcribed baculovirus gene PE-38. *Journal of Virology.* **65:** 805-812

S. Kurtz, Choudhuri J. V., Ohlebusch E., Schleiermacher C., Stoye J. and Giegerich R. (2001). REPuter: the manifold applications of repeat analysis on a genomic scale. *Nucleic Acids Research.* **29:** 4633-4642

J. Kuzio, Jaques R. and Faulkner P. (1989). Identification of p74, a gene essential for virulence of baculovirus occlusion bodies. *Virology.* **173:** 759-763

J. Kuzio, Pearson M. N., Harwood S. H., Funk C. J., Evans J. T., Slavicek J. M. and Rohrmann G. F. (1999). Sequence and analysis of the genome of a baculovirus pathogenic for Lymantria dispar. *Virology.* **253:** 17-34

M. Lange and Jehle J. A. (2003). The genome of the Cryptophlebia leucotreta granulovirus. *Virology.* **317:** 220-236

R. Lapointe, Back D. W., Ding Q. and Carstens E. B. (2000). Identification and molecular characterization of the Choristoneura fumiferana multicapsid nucleopolyhedrovirus genomic region encoding the regulatory genes pkip, p47, lef-12, and gta. *Virology.* **271:** 109-121

H. A. Lauzon, Lucarotti C. J., Krell P. J., Feng Q., Retnakaran A. and Arif B. M. (2004). Sequence and organization of the Neodiprion lecontei nucleopolyhedrovirus genome. *Journal of Virology.* **78:** 7023-7035

D. J. Leisy, Rasmussen C., Kim H. T. and Rohrmann G. F. (1995). The Autographa californica nuclear polyhedrosis virus homologous region 1a: identical sequences are essential for DNA replication activity and transcriptional enhancer function. *Virology.* **208:** 742-752

S. Y. Li. (2005). Virulence of a nucleopolyhedrovirus to Neodiprion abietis (Hymenoptera: Diprionidae). *Journal of Economic Entomology.* **98:** 1870-1875

Y. Li, Passarelli A. L. and Miller L. K. (1993). Identification, sequence, and transcriptional mapping of lef-3, a baculovirus gene involved in late and very late gene expression. *Journal of Virology.* **67:** 5260-5268

G. Lin and Blissard G. W. (2002). Analysis of an Autographa californica multicapsid nucleopolyhedrovirus lef-6-null virus: LEF-6 is not essential for viral replication but appears to accelerate late gene transcription. *Journal of Virology.* **76:** 5503-5514

G. Lin, Slack J. M. and Blissard G. W. (2001). Expression and localization of LEF-11 in Autographa californica nucleopolyhedrovirus-infected Sf9 cells. *The Journal of General Virology.* **82:** 2289-2294

F. Liu and Green M. R. (1990). A specific member of the ATF transcription factor family can mediate transcription activation by the adenovirus E1a protein. *Cell.* **61:** 1217-1224

H. S. Liu and Bilimoria S. L. (1990). Infected cell specific protein and viral DNA synthesis in productive and abortive infections of Spodoptera frugiperda nuclear polyhedrosis virus. *Archives of Virology.* **115:** 101-113

J. C. Liu and Maruniak J. E. (1995). Nucleotide sequence and transcriptional analysis of the gp41 gene of Spodoptera frugiperda nuclear polyhedrosis virus. *The Journal of General Virology.* **76 (Pt 6):** 1443-1450

J. J. Liu and Carstens E. B. (1995). Identification, localization, transcription, and sequence analysis of the Choristoneura fumiferana nuclear polyhedrosis virus DNA polymerase gene. *Virology.* **209:** 538-549

K. L. Lorick, Jensen J. P., Fang S., Ong A. M., Hatakeyama S. and Weissman A. M. (1999). RING fingers mediate ubiquitin-conjugating enzyme (E2)-dependent ubiquitination. *Proceedings of the National Academy of Sciences of the United States of America.* **96:** 11364-11369

A. Lu and Carstens E. B. (1992). Nucleotide sequence and transcriptional analysis of the p80 gene of Autographa californica nuclear polyhedrosis virus: a homologue of the Orgyia pseudotsugata nuclear polyhedrosis virus capsid-associated gene. *Virology.* **190:** 201-209

A. Lu and Carstens E. B. (1993). Immediate-early baculovirus genes transactivate the p143 gene promoter of Autographa californica nuclear polyhedrosis virus. *Virology.* **195:** 710-718

A. Lu, Craig A., Casselman R. and Carstens E. B. (1996). Nucleotide sequence, insertional mutagenesis, and transcriptional mapping of a conserved region of the baculovirus Autographa californica nuclear polyhedrosis virus (map unit 64.8-66.9). *Canadian Journal of Microbiology.* **42:** 1267-1273

A. Lu and Miller L. K. (1994). Identification of three late expression factor genes within the 33.8- to 43.4-map-unit region of Autographa californica nuclear polyhedrosis virus. *Journal of Virology.* **68:** 6710-6718

A. Lu and Miller L. K. (1995). The roles of eighteen baculovirus late expression factor genes in transcription and DNA replication. *Journal of Virology.* **69:** 975-982

M. Lundin, Nehlin J. O. and Ronne H. (1994). Importance of a flanking AT-rich region in target site recognition by the GC box-binding zinc finger protein MIG1. *Molecular and Cellular Biology.* **14:** 1979-1985

L. A. Lyons. (1964). The European pine sawfly, *Neodiprion sertifer* (Geoffr.) (Hymenoptera: Diprionidae). A review with emphasis on studies in Ontario. *Proceeds of the Entomological Society of Ontario.* **94:** 5-37

J. P. Mackay and Crossley M. (1998). Zinc fingers are sticking together. *Trends in Biochemical Sciences.* **23:** 1-4

J. P. Mackay, Kowalski K., Fox A. H., Czolij R., King G. F. and Crossley M. (1998). Involvement of the N-finger in the self-association of GATA-1. *Journal of Biological Chemistry.* **273:** 30560-30567

S. Maeda. (1989). Increased insecticidal effect by a recombinant baculovirus carrying a synthetic diuretic hormone gene. *Biochemical and Biophysical Research Communications.* **165:** 1177-1183

S. Maeda, Volrath S. L., Hanzlik T. N., Harper S. A., Majima K., Maddox D. W., Hammock B. D. and Fowler E. (1991). Insecticidal effects of an insect-specific neurotoxin expressed by a recombinant baculovirus. *Virology.* **184:** 777-780

T. H. Mainprize, Lee K. and Miller L. K. (1986). Variation in the temporal expression of overlapping baculovirus transcripts. *Virus Research.* **6:** 85-99

D. Mainz, Quadt I. and Knebel-Morsdorf D. (2002). Nuclear IE2 structures are related to viral DNA replication sites during baculovirus infection. *Journal of Virology.* **76:** 5198-5207

R. M. Mans and Knebel-Morsdorf D. (1998). In vitro transcription of pe38/polyhedrin hybrid promoters reveals sequences essential for recognition by the baculovirus-induced RNA polymerase and for the strength of very late viral promoters. *Journal of Virology.* **72:** 2991-2998

A. Marchler-Bauer, Anderson J. B., Cherukuri P. F., DeWeese-Scott C., Geer L. Y., Gwadz M., He S., Hurwitz D. I., Jackson J. D., Ke Z., Lanczycki C. J., Liebert C. A., Liu C., Lu F., Marchler G. H., Mullokandov M., Shoemaker B. A., Simonyan V., Song J. S., Thiessen P. A., Yamashita R. A., Yin J. J., Zhang D. and Bryant S. H. (2005). CDD: a

Conserved Domain Database for protein classification. *Nucleic Acids Research.* **33:** D192-6

M. E. Martignoni and P. J. Iwai. (1981). A catalogue of viral diseases ofinsects, mites, and ticks,other methods), 4th ed. Sinauer Associates, Sunderland, Mass. (H. D. Burges.) in Microbial control of pests and plant diseases 1970–1980. 1st ed., Academic Press, Inc., London, UK.

A. Martins and Shuman S. (2001). Mutational analysis of baculovirus capping enzyme Lef4 delineates an autonomous triphosphatase domain and structural determinants of divalent cation specificity. *Journal of Biological Chemistry.* **276:** 45522-45529

A. Martins and Shuman S. (2003). Mapping the triphosphatase active site of baculovirus mRNA capping enzyme LEF4 and evidence for a two-metal mechanism. *Nucleic Acids Research.* **31:** 1455-1463

V. Matys, Fricke E., Geffers R., Gossling E., Haubrock M., Hehl R., Hornischer K., Karas D., Kel A. E., Kel-Margoulis O. V., Kloos D. U., Land S., Lewicki-Potapov B., Michael H., Munch R., Reuter I., Rotert S., Saxel H., Scheer M., Thiele S. and Wingender E. (2003). TRANSFAC: transcriptional regulation, from patterns to profiles. *Nucleic Acids Research.* **31:** 374-378

B. F. McCutchen, Choudary P. V., Crenshaw R., Maddox D., Kamita S. G., Palekar N., Volrath S., Fowler E., Hammock B. D. and Maeda S. (1991). Development of a recombinant baculovirus expressing an insect-selective neurotoxin: potential for pest control. *Bio/technology (Nature Publishing Company).* **9:** 848-852

J. R. McLachlin and Miller L. K. (1994). Identification and characterization of vlf-1, a baculovirus gene involved in very late gene expression. *Journal of Virology.* **68:** 7746-7756

A. H. Meijer, de Kam R. J., d'Erfurth I., Shen W. and Hoge J. H. (2000). HD-Zip proteins of families I and II from rice: interactions and functional properties. *Molecular & General Genetics : MGG.* **263:** 12-21

H. Mi, Lazareva-Ulitsky B., Loo R., Kejariwal A., Vandergriff J., Rabkin S., Guo N., Muruganujan A., Doremieux O., Campbell M. J., Kitano H. and Thomas P. D. (2005). The PANTHER database of protein families, subfamilies, functions and pathways. *Nucleic Acids Research.* **33:** D284-8

L. K. Miller. (1999). An exegesis of IAPs: salvation and surprises from BIR motifs. *Trends in Cell Biology.* **9:** 323-328

L. K. Miller, Trimarchi R., Browne D. and Penncock G. D. A. (1983). A temperature-sensitive mutant of the baculovirus Autographa californica. *Virology.* **126:** 376-380

T. A. Mistretta and Guarino L. A. (2005). Transcriptional activity of baculovirus very late factor 1. *Journal of Virology.* **79:** 1958-1960

K. Mizugishi, Aruga J., Nakata K. and Mikoshiba K. (2001). Molecular properties of Zic proteins as transcriptional regulators and their relationship to GLI proteins. *Journal of Biological Chemistry.* **276:** 2180-2188

S. A. Monsma and Blissard G. W. (1995). Identification of a membrane fusion domain and an oligomerization domain in the baculovirus GP64 envelope fusion protein. *Journal of Virology.* **69:** 2583-2595

G. Moreau, Lucarotti C. J., Kettela E. G., Thurston G. S., Holmes S., Weaver C., Levin D. B. and Morin B. (2005). Aerial application of nucleopolyhedrovirus induces decline in increasing and peaking populations of Neodiprion abietis. *Biological Control.* **33:** 65-73

G. Moreau. (2006). Past and present outbreaks of the balsam fir sawfly in western Newfoundland: An analytical review. *Forest Ecology and Management.* , in press

T. D. Morris and Miller L. K. (1994). Mutational analysis of a baculovirus major late promoter. *Gene.* **140:** 147-153

B. Moser, Becnel J., White S., Afonso C., Kutish G., Shanker S. and Almira E. (2001). Morphological and molecular evidence that Culex nigripalpus baculovirus is an unusual member of the family Baculoviridae. *The Journal of General Virology.* **82:** 283-297

R. Muller, Pearson M. N., Russell R. L. and Rohrmann G. F. (1990). A capsid-associated protein of the multicapsid nuclear polyhedrosis virus of Orgyia pseudotsugata: genetic location, sequence, transcriptional mapping, and immunocytochemical characterization. *Virology.* **176:** 133-144

A. Muto and Osawa S. (1987). The guanine and cytosine content of genomic DNA and bacterial evolution. *Proceedings of the National Academy of Sciences of the United States*

of America. **84:** 166-169

M. S. Nissen and Friesen P. D. (1989). Molecular analysis of the transcriptional regulatory region of an early baculovirus gene. *Journal of Virology.* **63:** 493-503

H. Ochman, Lerat E. and Daubin V. (2005). Examining bacterial species under the specter of gene transfer and exchange. *Proceedings of the National Academy of Sciences of the United States of America.* **102:** 6595-6599

U. Ohler, Liao G. C., Niemann H. and Rubin G. M. (2002). Computational analysis of core promoters in the Drosophila genome. *Genome biology.* **3:** RESEARCH0087

M. Ohresser, Morin N., Cerutti M. and Delsert C. (1994). Temporal regulation of a complex and unconventional promoter by viral products. *Journal of Virology.* **68:** 2589-2597

E. Olofsson. (1988). Environmental persistence of the nuclear-polyhedrosis virus of the European pine sawfly in relation to epizootics in Swedish pine forests. *Journal of Invertebrate Pathology.* **52:** 119-129

E. Olofsson. (1973). Evaluation of a nuclear polyhedrosis virus as an agent for the control of the balsam fir sawfly, *Neodiprion abietis* Harr. Insect Pathology Research Institute, Department of the Environment, Canadian Forestry Service, Sault Ste. Marie, Ontario, Information Report IP-X-2

J. Olszewski and Miller L. K. (1997). Identification and characterization of a baculovirus structural protein, VP1054, required for nucleocapsid formation. *Journal of Virology.* **71:** 5040-5050

B. G. Ooi, Rankin C. and Miller L. K. (1989). Downstream sequences augment transcription from the essential initiation site of a baculovirus polyhedrin gene. *Journal of Molecular Biology.* **210:** 721-736

S. Osawa, Jukes T. H., Watanabe K. and Muto A. (1992). Recent evidence for evolution of the genetic code. *Microbiological reviews.* **56:** 229-264

A. L. Passarelli and Miller L. K. (1993a). Identification and characterization of lef-1, a baculovirus gene involved in late and very late gene expression. *Journal of Virology.* **67:**

3481-3488

A. L. Passarelli and Miller L. K. (1993b). Three baculovirus genes involved in late and very late gene expression: ie-1, ie-n, and lef-2. *Journal of Virology.* **67:** 2149-2158

A. L. Passarelli and Miller L. K. (1994). Identification and transcriptional regulation of the baculovirus lef-6 gene. *Journal of Virology.* **68:** 4458-4467

A. L. Passarelli, Todd J. W. and Miller L. K. (1994). A baculovirus gene involved in late gene expression predicts a large polypeptide with a conserved motif of RNA polymerases. *Journal of Virology.* **68:** 4673-4678

M. Pearson, Bjornson R., Pearson G. and Rohrmann G. (1992). The Autographa californica baculovirus genome: evidence for multiple replication origins. *Science.* **257:** 1382-1384

D. Q. Pham and Sivasubramanian N. (1992). In vivo transcriptional analysis of three baculovirus genes: evidence of homology between viral and host transcripts. *Virology.* **190:** 288-297

H. Piene, Ostaff D. and Eveleigh E. (2001). Growth loss and recovery following defoliation by the balsam fir sawfly in young, spaced balsam fir stands. *Canadian Entomologist.* **133:** 675-686

R. D. Possee and Howard S. C. (1987). Analysis of the polyhedrin gene promoter of the Autographa californica nuclear polyhedrosis virus. *Nucleic Acids Research.* **15:** 10233-10248

E. A. Prikhod'ko and Miller L. K. (1998). Role of baculovirus IE2 and its RING finger in cell cycle arrest. *Journal of Virology.* **72:** 684-692

E. A. Prikhod'ko and Miller L. K. (1999). The baculovirus PE38 protein augments apoptosis induced by transactivator IE1. *Journal of Virology.* **73:** 6691-6699

S. S. Pullen and Friesen P. D. (1995a). The CAGT motif functions as an initiator element during early transcription of the baculovirus transregulator ie-1. *Journal of Virology.* **69:** 3575-3583

S. S. Pullen and Friesen P. D. (1995b). Early transcription of the ie-1 transregulator gene of Autographa californica nuclear polyhedrosis virus is regulated by DNA sequences within its 5' noncoding leader region. *Journal of Virology.* **69:** 156-165

I. Quadt, Mainz D., Mans R., Kremer A. and Knebel-Morsdorf D. (2002). Baculovirus infection raises the level of TATA-binding protein that colocalizes with viral DNA replication sites. *Journal of Virology.* **76:** 11123-11127

J. C. Rapp, Wilson J. A. and Miller L. K. (1998). Nineteen baculovirus open reading frames, including LEF-12, support late gene expression. *Journal of Virology.* **72:** 10197-10206

K. K. Rashidan, Nassoury N., Giannopoulos P. N. and Guertin C. (2005). Transcription, translation, and immunolocalization of ODVP-6E/ODV-E56 and p74 proteins: two highly conserved ODV-associated envelope proteins of Choristoneura fumiferana Granulovirus. *Journal of Biochemistry and Molecular Biology.* **38:** 65-70

F. B. Razuck, Ribeiro B., Vargas J. H., Wolff J. L. and Ribeiro B. M. (2002). Characterization of the p10 gene region of Anticarsia gemmatalis nucleopolyhedrovirus. *Virus Genes.* **24:** 243-247

B. M. Ribeiro, Hutchinson K. and Miller L. K. (1994). A mutant baculovirus with a temperature-sensitive IE-1 transregulatory protein. *Journal of Virology.* **68:** 1075-1084

T. E. Roberts and Faulkner P. (1989). Fatty acid acylation of the 67K envelope glycoprotein of a baculovirus: Autographa californica nuclear polyhedrosis virus. *Virology.* **172:** 377-381

S. M. Rodems and Friesen P. D. (1993). The hr5 transcriptional enhancer stimulates early expression from the Autographa californica nuclear polyhedrosis virus genome but is not required for virus replication. *Journal of Virology.* **67:** 5776-5785

D. Z. Rohel and Faulkner P. (1984). Time Course Analysis and Mapping of *Autographa californica* Nuclear Polyhedrosis Virus Transcripts. *Journal of Virology.* **50:** 739-747

M. Rosinski, Reid S. and Nielsen L. K. (2002). Kinetics of baculovirus replication and release using real-time quantitative polymerase chain reaction. *Biotechnology and Bioengineering.* **77:** 476-480

L. D. Rothman and Myers J. H. (1996). Debilitating effects of viral diseases on host Lepidoptera. *Journal of Invertebrate Pathology.* **67:** 1-10

R. L. Russell, Pearson M. N. and Rohrmann G. F. (1991). Immunoelectron microscopic examination of Orgyia pseudotsugata multicapsid nuclear polyhedrosis virus-infected Lymantria dispar cells: time course and localization of major polyhedron-associated proteins. *The Journal of General Virology.* **72 (Pt 2):** 275-283

R. L. Russell and Rohrmann G. F. (1993). A 25-kDa protein is associated with the envelopes of occluded baculovirus virions. *Virology.* **195:** 532-540

R. L. Russell and Rohrmann G. F. (1997). Characterization of P91, a protein associated with virions of an Orgyia pseudotsugata baculovirus. *Virology.* **233:** 210-223

G. S. Salvesen and Duckett C. S. (2002). IAP proteins: blocking the road to death's door. *Nature reviews. Molecular Cell Biology.* **3:** 401-410

R. Schuh, Aicher W., Gaul U., Cote S., Preiss A., Maier D., Seifert E., Nauber U., Schroder C. and Kemler R. (1986). A conserved family of nuclear proteins containing structural elements of the finger protein encoded by Kruppel, a Drosophila segmentation gene. *Cell.* **47:** 1025-1032

J. Schultz, Copley R. R., Doerks T., Ponting C. P. and Bork P. (2000). SMART: a web-based tool for the study of genetically mobile domains. *Nucleic Acids Research.* **28:** 231-234

S. Schwartz, Zhang Z., Frazer K. A., Smit A., Riemer C., Bouck J., Gibbs R., Hardison R. and Miller W. (2000). PipMaker--a web server for aligning two genomic DNA sequences. *Genome Research.* **10:** 577-586

B. S. Shastry. (1996). Transcription factor IIIA (TFIIIA) in the second decade. *Journal of Cell Science.* **109:** 535-539

W. F. Shen, Montgomery J. C., Rozenfeld S., Moskow J. J., Lawrence H. J., Buchberg A. M. and Largman C. (1997). AbdB-like Hox proteins stabilize DNA binding by the Meis1 homeodomain proteins. *Molecular and Cellular Biology.* **17:** 6448-6458

T. R. Shieh. (1989). Industrial production of viral pesticides. *Advances in Virus Research.* **36:** 315-343

O. Simon, Gutierrez S., Williams T., Caballero P. and Lopez-Ferber M. (2005). Nucleotide sequence and transcriptional analysis of the pif gene of Spodoptera frugiperda nucleopolyhedrovirus (SfMNPV). *Virus Research.* **108:** 213-220

W. A. Smirnoff. (1972). The effect of sunlight on the nuclear-polyhedrosis virus of Neodiprion swainei with measurement of the solar energy recieved. *Journal of Invertebrate Pathology.* **19:** 179-188

G. E. Smith, Vlak J. M. and Summers M. D. (1982). In Vitro Translation of *Autographa californica* Nuclear Polyhedrosis Virus Early and Late mRNAs. *Journal of Virology.* **44:** 199-208

S. Sriram and Gopinathan K. P. (1998). The potential role of a late gene expression factor, lef2, from Bombyx mori nuclear polyhedrosis virus in very late gene transcription and DNA replication. *Virology.* **251:** 108-122

E. A. Steinhaus. (1948). Polyhedrosis, ("wilt disease") of the alfalfa caterpillar. *Journal of Economic Entomology.* **41:** 859-865

L. M. Stewart, Hirst M., Lopez Ferber M., Merryweather A. T., Cayley P. J. and Possee R. D. (1991). Construction of an improved baculovirus insecticide containing an insect-specific toxin gene. *Nature.* **352:** 85-88

D. L. Swofford. PAUP*: Phylogenetic Analysis Using Parsimony (and Other Methods), Version 4.0. *Sunderland, MA: Sinauer Assoc.*

D. A. Theilmann and Stewart S. (1991). Identification and characterization of the IE-1 gene of Orgyia pseudotsugata multicapsid nuclear polyhedrosis virus. *Virology.* **180:** 492-508

D. A. Theilmann and Stewart S. (1992). Molecular analysis of the trans-activating IE-2 gene of Orgyia pseudotsugata multicapsid nuclear polyhedrosis virus. *Virology.* **187:** 84-96

S. M. Thiem and Miller L. K. (1989). Identification, sequence, and transcriptional mapping of the major capsid protein gene of the baculovirus Autographa californica nuclear polyhedrosis virus. *Journal of Virology.* **63:** 2008-2018

E. D. Thomas, Reichelderfer C. F. and Heimpel A. M. (1973). The effect of soil pH on the persistence of cabbage looper nuclear polyhedrosis virus in soil. *Journal of Invertebrate Pathology.* **21:** 21-25

C. G. Thompson and Scott D. W. (1979). Production and persistence of the nuclear polyhedrosis virus of the Douglas-fir Tussak moth, *Orgyia pseudotsugata* (Lepidoptera: Lymantridae) in the forest ecosystem. *Journal for Invertebrate Pathology.* **33:** 57-65

J. D. Thompson, Higgins D. G. and Gibson T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research.* **22:** 4673-4680

J. W. Todd, Passarelli A. L., Lu A. and Miller L. K. (1996). Factors regulating baculovirus late and very late gene expression in transient-expression assays. *Journal of Virology.* **70:** 2307-2317

J. W. Todd, Passarelli A. L. and Miller L. K. (1995). Eighteen baculovirus genes, including lef-11, p35, 39K, and p47, support late gene expression. *Journal of Virology.* **69:** 968-974

M. D. Tomalski, Wu J. G. and Miller L. K. (1988). The location, sequence, transcription, and regulation of a baculovirus DNA polymerase gene. *Virology.* **167:** 591-600

E. A. van Strien, Ijkel W. F., Gerrits H., Vlak J. M. and Zuidema D. (2000). Characteristics of the transactivator gene ie1 of Spodoptera exigua multiple nucleopolyhedrovirus. *Archives of Virology.* **145:** 2115-2133

E. A. van Strien, Zuidema D., Goldbach R. W. and Vlak J. M. (1992). Nucleotide sequence and transcriptional analysis of the polyhedrin gene of Spodoptera exigua nuclear polyhedrosis virus. *The Journal of General Virology.* **73 (Pt 11):** 2813-2821

P. Viswanathan, Venkaiah B., Kumar M. S., Rasheedi S., Vrati S., Bashyam M. D. and Hasnain S. E. (2003). The homologous region sequence (hr1) of Autographa californica multinucleocapsid polyhedrosis virus can enhance transcription from non-baculoviral

promoters in mammalian cells. *Journal of Biological Chemistry.* **278:** 52564-52571

L. E. Volkman, Goldsmith P. A., Hess R. T. and Faulkner P. (1984). Neutralization of budded Autographa californica NPV by a monoclonal antibody: identification of the target antigen. *Virology.* **133:** 354-362

D. R. Wallace and J. C. Cunningham. (1995). Diprionid Sawflies (J. A. Armstrong and W. G. H. Ives.) in Forest Insect Pests in Canada. 1st ed., Natural Resources Canada, Ottawa,.

H. Wang, Chen X., Wang H., Arif B. M., Vlak J. M. and Hu Z. (2001). Nucleotide sequence and transcriptional analysis of a putative basic DNA-binding protein of Helicoverpa armigera nucleopolyhedrovirus. *Virus genes.* **22:** 113-120

P. Wang and Granados R. R. (2001). Molecular structure of the peritrophic membrane (PM): identification of potential PM target sites for insect control. *Archives of Insect Biochemistry and Physiology.* **47:** 110-118

W. Wang, Davison S. and Krell P. J. (2004). Identification and characterization of a major early-transcribed gene of Trichoplusia ni single nucleocapsid nucleopolyhedrovirus using the baculovirus expression system. *Virus genes.* **29:** 19-29

J. O. Washburn, Kirkpatrick B. A. and Volkman L. E. (1995). Comparative pathogenesis of Autographa californica M nuclear polyhedrosis virus in larvae of Trichoplusia ni and Heliothis virescens. *Virology.* **209:** 561-568

A. M. Weissman. (2001). Themes and variations on ubiquitylation. *Nature reviews. Molecular Cell Biology.* **2:** 169-178

U. Weyer and Possee R. D. (1988). Functional analysis of the p10 gene 5' leader sequence of the Autographa californica nuclear polyhedrosis virus. *Nucleic Acids Research.* **16:** 3635-3653

M. Whitford and Faulkner P. (1992). Nucleotide sequence and transcriptional analysis of a gene encoding gp41, a structural glycoprotein of the baculovirus Autographa californica nuclear polyhedrosis virus. *Journal of Virology.* **66:** 4763-4768

M. E. Wilson and Consigli R. A. (1985). Functions of a protein kinase activity associated with purified capsids of the granulosis virus infecting *plodia interpunctella*. *Virology.* **143:** 526-535

E. Wingender. (1988). Compilation of transcription regulating proteins. *Nucleic Acids Research.* **16:** 1879-1902

S. A. Wolfe, Nekludova L. and Pabo C. O. (2000). DNA recognition by Cys2His2 zinc finger proteins. *Annual Review of Biophysics and Biomolecular Structure.* **29:** 183-212

G. M. Wolgamot, Gross C. H., Russell R. L. and Rohrmann G. F. (1993). Immunocytochemical characterization of p24, a baculovirus capsid-associated protein. *The Journal of General Virology.* **74:** 103-107

S. Wormleaton, Kuzio J. and Winstanley D. (2003). The complete sequence of the Adoxophyes orana granulovirus genome. *Virology.* **311:** 350-365

C. H. Wu, Huang H., Yeh L. S. and Barker W. C. (2003). Protein family classification and functional annotation. *Computational Biology and Chemistry.* **27:** 37-47

C. T. Wu and Morris J. R. (1999). Transvection and other homology effects. *Current Opinion in Genetics & Development.* **9:** 237-246

X. Wu, Stewart S. and Theilmann D. A. (1993a). Alternative transcriptional initiation as a novel mechanism for regulating expression of a baculovirus trans activator. *Journal of Virology.* **67:** 5833-5842

X. Wu, Stewart S. and Theilmann D. A. (1993b). Characterization of an early gene coding for a highly basic 8.9K protein from the Orgyia pseudotsugata multicapsid nuclear polyhedrosis virus. *The Journal of General Virology.* **74:** 1591-1598

F. Yang, He J., Lin X., Li Q., Pan D., Zhang X. and Xu X. (2001). Complete genome sequence of the shrimp white spot bacilliform virus. *Journal of Virology.* **75:** 11811-11820

S. Yang and Miller L. K. (1998a). Control of baculovirus polyhedrin gene expression by very late factor 1. *Virology.* **248:** 131-138

S. Yang and Miller L. K. (1998b). Expression and mutational analysis of the baculovirus very late factor 1 (vlf-1) gene. *Virology.* **245:** 99-109

S. Yang and Miller L. K. (1999). Activation of baculovirus very late promoters by interaction with very late factor 1. *Journal of Virology.* **73:** 3404-3409

G. S. Yochum and Ayer D. E. (2001). Pf1, a novel PHD zinc finger protein that links the TLE corepressor to the mSin3A-histone deacetylase complex. *Molecular and Cellular Biology.* **21:** 4110-4118

P. M. Zanotto, Kessing B. D. and Maruniak J. E. (1993). Phylogenetic interrelationships among baculoviruses: evolutionary rates and host associations. *Journal of Invertebrate Pathology.* **62:** 147-164

E. M. Zdobnov and Apweiler R. (2001). InterProScan--an integration platform for the signature-recognition methods in InterPro. *Bioinformatics (Oxford, England).* **17:** 847-848

E. A. Zemskov, Kang W. and Maeda S. (2000). Evidence for nucleic acid binding ability and nucleosome association of Bombyx mori nucleopolyhedrovirus BRO proteins. *Journal of Virology.* **74:** 6784-6789

C. X. Zhang, Ma X. C. and Guo Z. J. (2005). Comparison of the complete genome sequence between C1 and G4 isolates of the Helicoverpa armigera single nucleocapsid nucleopolyhedrovirus. *Virology.* **333:** 190-199