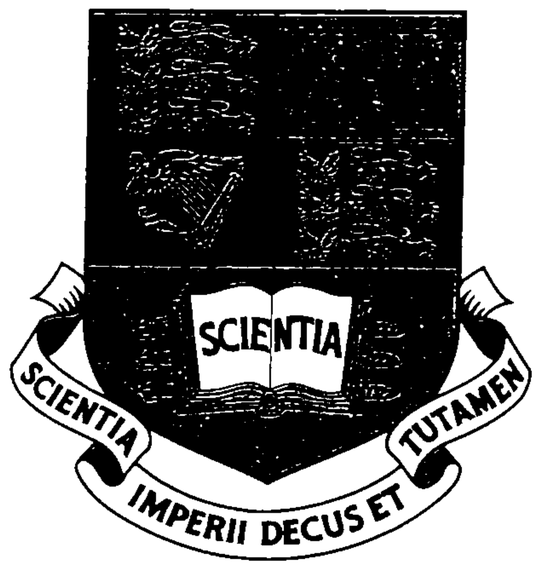
Imperial College
London



Use of Comparative Genomics and Phylogenetics to Study the Evolution of the *Baculoviridae*

Elisabeth A. Herniou

Thesis submitted for the degree of
Doctor of Philosophy of the University of London
February 2003



Department of Biological Sciences - Imperial College London
Ecology and Biocontrol Group - NERC CEH-Oxford



DECLARATION

This declaration is to confirm that the work presented in this thesis is my own.

Chapter 2 is based on two publications. The first was published in 2001 in *Journal of Virology* with Teresa Luque, Xinwen Chen, Just Vlak, Doreen Winstanley, Jenny Cory, and David O'Reilly. The second was published in the 2003 edition of *Annual Review of Entomology* with Julie Olszewski, Jenny Cory and David O'Reilly.

Chapter 3 will be the object of several publications in collaboration with Julie Olszewski, Jenny Cory and David O'Reilly.

I collected and analysed all the data reported in Chapters 2 & 3 and wrote the manuscript with advice and comments from my co-authors.

PhD candidate:

Elisabeth Herniou

Supervisors:

Dr David O'Reilly

Dr Julie Olszewski

Dr Jenny Cory

ABSTRACT

This thesis investigates evolutionary relationships between viruses of the family *Baculoviridae*. I have been taken two approaches to study different aspects of baculovirus evolution. The first was to compare complete genomes in an evolutionary context and the second to explore the phylogenetic diversity of the whole family.

The complete genome sequence determination of 13 baculoviruses has been a major resource for enhancing the understanding of their biology. I have applied and developed new methods to use complete genome sequences for the reconstruction of baculovirus phylogenies based on gene sequences, gene order and gene content. The history of gene acquisitions and losses was visualised by mapping the gene content onto the baculovirus phylogenetic tree. This highlighted key genetic innovations that have accompanied baculovirus diversification. Furthermore, the use of comparative genomics methods has shown functional linkage between a number of genes, revealing 2 putative DNA repair systems, Lymantriid host related factors and a *Baculoviridae* core gene cluster. Thus, unravelling genomic evolutionary history can be useful for understanding the biology of baculoviruses and how it is linked to their hosts.

Baculoviruses are primarily insect pathogens, infecting mostly larvae of the order Lepidoptera but also Diptera and Hymenoptera. I have explored the diversity of the *Baculoviridae* and in particular how this related to their hosts. From field isolates of infected insects of the historical virus collection of NERC-CEH-Oxford, I acquired sequences of the *polyhedrin, lef-8* and *ac22* genes. These were employed to reconstruct molecular phylogenies to describe baculovirus evolutionary relationships. The trees were used to investigate the relationship between virus isolates and virus species and to reconstruct ancestral host use. The host of the ancestral lepidopteran baculovirus was found to have belonged to the family Noctuidae. Furthermore, the phylogenetic analyses showed that the evolution of baculovirus host affiliation is characterised by phylogenetic conservatism. This is reflected both at the insect order level, where baculoviruses from Lepidoptera, Hymenoptera and Diptera cluster separately, and at the family level within the Lepidoptera. Thus, surveying the diversity of the *Baculoviridae* clearly highlighted the evolutionary link between baculoviruses and their hosts.

ACKNOWLEDGEMENTS

First of all, I would like to acknowledge my three supervisors, Jenny Cory, Julie Olszewski and David O'Reilly. I owe to David and Jenny to have entrusted me with this PhD project. I am very grateful to Julie for jumping on board, for her constant support and most of all for always keeping her door open. I would like to thank all three of them for thought provoking discussions and for their guidance throughout my PhD.

My research has been supported by the Natural Environment Research Council through the funding of a CASE studentship, award GT04/99/TS/142. I am thankful to the Society for Invertebrate Pathology, Society of General Microbiology and the Gregory Fund to have founded travel awards to present my work at the IUMS Congress and at the SIP meeting in 2002. I am also grateful to Imperial College for awarding me the 2001 Marshall Scholarship for research in biology.

For fruitful discussions on phylogenetic analyses, I would like to thank my advisers M. Tristem and A. Burt, as well as David Posada, Tim Littlewood, Mark Wilkinson and Carlos López Vaamonde.

I am most thankful to Renée Lapointe, Teresa Luque, Isabelle Nobiron, Lluísa Vilaplana, Eloisa Arana and Yeon Ho Je for discussions on baculoviruses and most of all for their cheerful support during the last three years.

I thank my parents for initially giving me the courage to move to the UK to undertake a PhD and for never doubting of the outcome. Lastly, my deepest thanks have to go to Carlos because he was always there, encouraging and supportive. Without him, none of this would have been possible.

TABLE OF CONTENT

DECLARATION 2

ABSTRACT.....3

ACKNOWLEDGMENTS 4

TABLE OF CONTENT.....5

LIST OF TABLES 7

LIST OF FIGURES.....8

LIST OF APPENDICES 9

I. CHAPTER ONE: INTRODUCTION.....10

A. General introduction to the baculoviruses.....11

1. Historical perspective: A virus on the "Silk Road" trail 11

2. Taxonomy.....14

3. Life cycle - Pathogenesis.....15

4. Modes of evolution 19

5. Phylogeny.....22

B. Thesis Outline.....25

1. Chapter 2: use of complete genomes to investigate baculovirus evolution 25

2. Chapter 3: Sampling Baculovirus Diversity and Evolution 25

II. CHAPTER TWO: THE USE OF COMPLETE GENOMES TO INVESTIGATE BACULOVIRUS EVOLUTION 27

A. Baculovirus Genomes 28

1. Core Baculovirus Genes.....29

2. Lepidopteran vs Dipteran Baculoviruses. 32

B. Use Of Whole Genome Sequence Data To Infer Baculovirus Phylogeny.....34

1. Methods Of Phylogenetic Reconstruction Based On Complete Genomes 35

2. Results of the Phylogenetic Analyses Based On 9 Complete Genome Sequences.....42

3. Results of the Phylogenetic Analyses Based On 13 Complete Genome Sequences.....51

4. A Phylogeny Of The Baculoviridae.....55

C. Comparative genomics.....57

1. The evolution of genome content 57

2. Discovering functional linkage.....65

D. Summary 71
III. CHAPTER THREE: SAMPLING BACULOVIRUS DIVERSITY AND EVOLUTION 73
A. Baculovirus Diversity 74
1. Baculovirus Nomenclature 75
2. Baculovirus Phylogenies 78
B. Material and Methods 80
1. The baculovirus collection 80
2. Viral DNA Purification 80
3. Primer design 83
4. PCR amplification and sequencing 85
5. Phylogenetic analyses 86
6. Looking for extra genetic information 88
C. Position of non-lepidopteran baculoviruses within the *Baculoviridae* 91
1. Samples examined 91
2. PCR & Sequencing results 92
3. Topo Walking Results 96
4. Phylogenetic Analyses 98
5. Discussion 104
D. Phylogeny of Lepidopteran baculoviruses 109
1. Taxa and Sequence Sampling 109
2. Phylogenetic Analyses of the Polyhedrin Gene Fragment 113
3. Phylogenetic Analyses of the Lef-8 Gene Fragment 117
4. Clustering Viral Strains Into Species: Do the polh and lef-8 phylogenies help? 120
5. Measuring the incongruence between the polh and lef-8 phylogenies 129
6. Conclusions 132
E. Lepidopteran Hosts Shape Baculovirus Evolution 133
1. Introduction: Bases of Baculovirus-Host Co-Evolution. 133
2. Methods: Mapping Baculovirus Host Use 134
3. Results and Discussion 137
F. Summary 146
IV. REFERENCES 147
V. APPENDICES 168

LIST OF TABLES

Table 1: Genes present in all baculovirus genomes 29

Table 2: Complete baculovirus genome sequences 36

Table 3: Tree topologies comparisons – SH and KH tests 45

Table 4: Relative breakpoint distance matrix 48

Table 5: Details of the gene acquisitions and losses, which have occurred during baculovirus evolution 60

Table 6: Taxonomic structure of the baculovirus family 76

Table 7: Number of baculovirus sequences available in the database early in 2000.....83

Table 8: Design of degenrate oligonucleotide primers used for the amplification of baculoviruses.....84

Table 9: Viruses from non-lepidopteran insects examined in this study.....92

Table 10: Results of the similarity searches for the non-lepidopteran viral sequences ..94

Table 11: Summary of the experiments done and the results obtained for the Topo Walking protocol.....97

Table 12: Statisctical differences of the ML trees obtained for the Lef-8+Ac22, Lef-8 and Ac22 data sets.....100

Table 13: List of the lepidopteran baculoviruses used in the *polyhedrin* and *lef-8* phylogenetic analyses.....110

Table 14: Statisctical differences between the *polh* phylogenies obtained under different topological constraints.....116

Table 15: Classification of the lepidopteran hosts of the baculoviruses used in this study – Higher level classification of the Lepidoptera after Scoble (1992) 135

LIST OF FIGURES

Figure 1: Description of baculovirus infection symptoms in ancient silkworm litterature 12

Figure 2: Phenotypes of baculovirus occlusion bodies 15

Figure 3: Baculovirus infection cycle 16

Figure 4: Baculovirus replication at cellular level 18

Figure 5: Phenotypes of baculovirus virions 19

Figure 6: Baculovirus phylogeny 23

Figure 7: Circular map of the CpGV genome.....28

Figure 8: Gene data sets used in the complete genomes phylogenetic analyses 39

Figure 9: MPT topologies obtained from the phylogenetic analyses of the 63 genes shared by the 9 baculovirus genomes.....44

Figure 10: Phylogenies based on the complete genomes of 9 baculoviruses 46

Figure 11: *odv-e66* and *polyhedrin* gene phylogenies.....47

Figure 12: Phylogenies based on the complete genomes of 13 baculoviruses 52

Figure 13: Patterns of gene content evolution, possible mapping of gene acquisitions and losses on a phylogeny 58

Figure 14: Gene content mapping 59

Figure 15: Phylogenetic profiles 67

Figure 16: Schematic map of the *Baculoviridae* core gene cluster.....69

Figure 17: Gene parity plots 70

Figure 18: Degenerate primer design and strategy for the PCr and direct sequencing of baculovirus genes 85

Figure 19: Schematic description of the Topo Walker kit protocol.....89

Figure 20: ML trees obtained for the Lef-8+Ac22, Lef-8 and Ac22 data sets 102

Figure 21: Robustness of the Lef-8+Ac22 tree.....103

Figure 22: Evolution of the *Baculoviridae* 106

Figure 23: *Polyhedrin* phylogeny.....115

Figure 24: *Lef-8* phylogeny 119

Figure 25: List of lepidopteran *Nucleopolyhedrovirus* species 123

Figure 26: List of *Granulovirus* species.....127

Figure 27: Congruence between the polh and lef-8 trees 131

Figure 28: Reconstruction of ancestral host families on the *polyhedrin* phylogeny.....138

Figure 29: Reconstruction of ancestral host super-families on the *polyhedrin* phylogeny 142
Figure 30: Reconstruction of ancestral host families and super-families on the *lef-8* phylogeny 143

LIST OF APPENDICES

Appendix 1: Gene homologies relationships 169
Appendix 2: Virus collection 183
Appendix 3: *polyhedrin* and *lef-8* DNA sequence alignments 187

I. Chapter One:

Introduction

A. General introduction to the baculoviruses

1. Historical perspective: A virus on the "Silk Road" trail

Chinese legends attribute the discovery of sericulture, the production of silk, to the princess Si-Ling-Chi in the year 2700 B.C. (Cherry, 1993). From the start, silk was of great economic importance. Its commercialisation across Eurasia opened the "Silk Road" linking China to the Occident. The stakes were so high that Roman Emperors sent spies to steal the secret of silk production and in the process break the Chinese monopoly of the precious fabric. Around 550 A.D., Persian monks smuggled silkworm eggs and mulberry seeds to the Byzantine Empire. Since then sericulture has developed in the Mediterranean basin and across Europe.

Because the silkworm, *Bombyx mori* (L), has been domesticated for thousand of years, it is not surprising that pathogens outbreaks, including baculoviruses, have been observed historically. The first written description in the Occident of an insect disease caused by a baculovirus has been attributed to Marco Girolamo Vida (Benz, 1986). In 1527, the priest from Cremona (Italy) published "*De Bombyce*" (Vida, 1527), a Latin poem in which he describes the culture of the silkworm (figure 1A). In this text, some passages describing diseases of silkworms can be interpreted as classic symptoms of baculovirus infection. For example, "... *Protinus aegras delapse effundunt aminas, lecemque relinquunt.* ...", which translates into English as "... *Once they have fallen to the disease, the fluids escape as their life departs.* ..." can be interpreted as a description of the way infected caterpillar cadavers break down to disperse virus particles.

Additional historical evidence of baculovirus disease can be found in the French literature of the late 18th century when renewed interest in the silk industry led to the publication of a number of books on silkworms. In one such book (figure 1B), Reynaud, a silk producer from the Cevennes, describes a number of silkworm pathologies including a disease termed "grasserie" (Reynaud, 1824). Caterpillars affected by grasserie become fat, change colour, do not moult, and exude viscous liquid after their death. These symptoms correspond exactly to those of a baculovirus infection.

A.

MARCI HIERONYMI VIDAE CREMONEN. BOMBYCVM LIBER PRIMVS.

V os mores, quas aut parvis reptan-
tibus aries
I uppiter addiderit, quæ fila tenacia
Serum
O re vomunt salutare, nos mecum evolvite Nymphæ
S eriades. nos lanifici crepis aurea nostis
M unera, quæ pater Italiæ prior intulit olim
S erius huc patriis Serum devectus ab oris.
T ugue ades, et nostro succurre Isabella labori
N ympha Padi in ripa magnis e regibus orta,
Q uæ gentem pulchra auxisti pulcherrima prole
G onsagam, exiguis ades huc non aspera cœptis.
E t vos o placidæ properate, ubi quæque, puellæ,
V estrum opus hoc, vestros facites hic discite cultus.

G ii

B.

DES

VERS A SOIE;

ET DE LEUR ÉDUCATION

SELON LA PRATIQUE

DES CÉVENNES;

Suivi d'une Notice sur les divers produits de la Soie, et sur la manière de tirer les Bastides et les Filoselles; avec des Notions sur la fabrique des Etoffes de Gazzage;

PAR M. REYNAUD,



PARIS.

A LA LIBRAIRIE DU COMMERCE,

Chez Renard, rue Saint-Jean, n. 71.

1824.

(101)

Les vers affectés de cette maladie quittent la feuille, et vont périr çà et là sur le bord des ateliers, dans une espèce d'état de marasme et de langueur.

La maladie des gras ou grasserie ne commence guères à se déclarer que vers la seconde mue, et pendant le troisième âge suivant. Elle paraît provenir d'une nourriture trop substantielle pour cette époque, lorsqu'on n'a pas soin de fournir aux vers des feuilles d'une espèce tendre, ou cueillies sur de jeunes arbres. Les vers gras continuent de manger; ils grossissent ou enflent, mais leur corps est opaque et de couleur verdâtre, probablement à cause de la feuille qu'ils ne digèrent pas. L'humeur malfaisante qui se forme dans leur corps, transude au dehors, et elle se répand sous forme de liqueur visqueuse et sale. L'insecte

(102)

meurt, pour l'ordinaire, deux ou trois jours après qu'il a passé le temps de la mue, dans il ne subit point la crise(36).

Il paraît encore que dès le moment de la couvée, le ver à soie contracte des dispositions à cette maladie, lorsque la graine a été mal hivernée, que les vers éclosent spontanément, ou par l'artifice d'une chaleur animale, mais accompagnée de transpirations humides, sans que la graine ait été aérée suffisamment par une fréquente ouverture des nouets ou petits paquets où elle est enfermée.

On ne spécifie pas bien précisément en quoi consiste le vice des feuilles réputées capables de donner des gras. On dit que la feuille jaunie par la gelée peut contribuer au développement de cette maladie; on dit aussi, mais bien plus déraisonnablement, que les bourgeons donnés aux vers avancés la pé-

Figure 1: Description of Baculovirus Infection Symptoms in Ancient Silkworm Litterature. A: "*De Bombyce*", a poem by Vida (1527), 1st page. B: "*Des Vers à Soie*", a monograph by Reynaud (1824), 1st page and p101-102 describing the symptoms of *grasserie*.

However grasserie was not the only known, nor the most problematic disease of silkworms. Pebrine (microsporidium) and flacherie (picorna-like virus) were far more widespread in the insectaries. Their impact on silk production became so important in the 1850s that the French government appointed Louis Pasteur to characterise the pathogens and devise a protocol to alleviate the epizootics (Pasteur, 1870). The industry recovered but was again affected in the 1920s by another wave of epizootics caused by flacherie and this time, grasserie. André Paillot was commissioned to study the pathogens in 1923. Four years later, he described the grasserie causing agent, which was then classified as polyhedrosis disease (Paillot, 1927). At a time when cellular structures were still being identified, histopathology showed that grasserie caused cell destruction and that geometric crystals accumulated in the nucleus of infected cells (Paillot, 1927). This is still an appropriate description of the pathology of a nucleopolyhedrovirus.

In the 1940s, advances in microscopy lead to the first visualisation of the rod-shaped baculovirus virion (Bergold, 1953). At the same time, the use of baculoviruses as biological control agents was being developed (Balch & Bird, 1944). It is the use of baculoviruses as pesticides that has driven much of the interest in baculovirology since the 1950s. Baculoviruses have been characterised from a large number of arthropod hosts (Martignoni & Iwai, 1986) and used across the world against a variety of agricultural and forest pests (Moscardi, 1999). Technological advances during the 1970s-80s enabled the genetic improvement of baculoviruses and the development of baculovirus gene expression systems (Miller, 1997).

There is one last twist to the story that links the silk trail and baculoviruses. Around the time of Pasteur's studies on silkworms, the American silk industry was also experiencing difficulties due to widespread epizootics affecting *B. mori*. Attempts were made to improve caterpillar resistance to diseases. Pr L. Trouvelot conducted one such attempt in 1869. He intended to produce silk-forming hardy hybrids by crossing *B. mori* and *Lymantria dispar*. Needless to say he was unsuccessful. However, it is worth mentioning because a consequence of these experiments was the escape of the gypsy moth, *L. dispar* (Cherry, 1993). This species has since become one of the most serious and widespread North American forest pests. Baculoviruses are widely utilised for the control of gypsy moths in the United States and Canada (Moscardi, 1999), showing that what might have been disastrous to the silk industry could prove beneficial to the forest industry.

2. Taxonomy

a) Viral taxonomy

The taxonomy and nomenclature of baculoviruses are subject to rules set by the International Committee on Taxonomy of Viruses (ICTV). This committee was set up in the late 1960s, to respond to a desire to unify the way viruses are named and classified. The first level of virus classification is based on genome structure: DNA/RNA, double/single stranded, circular/linear, single/multiple fragments, size. The next levels are based on biology and infection cycles: host, cell specificity, in other terms ecological niche (van Regenmortel et al., 2000).

One of the major problems of viral taxonomy is related to the definition of a species. The definition for cellular organisms is one of the major questions in biology and one that has given rise to a number of concepts, such as biological species, based on populations (Mayr, 1982) and evolutionary species, based on phylogenetic lineages (Wiley, 1981). However these concepts are difficult to apply to viruses, and need to be used with care, depending on the information available for each virus species. This is rendered more challenging by the fact that viruses evolve extremely rapidly, to the point that it is sometimes difficult to recognise well-characterised laboratory strains in the wild, and conversely to identify field isolates. This gives rise to a multiplication of names that might not always be relevant in terms of the true virus species.

b) Definition of the Baculoviridae

The family *Baculoviridae* comprises a diverse group of arthropod-specific viruses. Baculoviruses are rod-shaped, enveloped viruses with circular double stranded DNA genomes ranging in size from 90 kb to 180 kb, and which initiate infection from an occluded form (Blissard et al., 2000). They have been reported worldwide from over 600 host species (Martignoni & Iwai, 1986), mostly from the order Lepidoptera but also from the orders Diptera, Hymenoptera and the crustacean order Decapoda (Blissard et al., 2000). The family *Baculoviridae* is currently subdivided into two genera, based largely on the morphology of the occlusion bodies (OBs) they form in infected cells (Blissard et al., 2000). The genus *Nucleopolyhedrovirus* (NPV) is characterized by viruses forming polyhedral OBs each containing many virions (Rohrmann, 1999), whereas viruses of the genus *Granulovirus* (GV) typically produce ovoid OBs with a single virion (Winstanley & O'Reilly, 1999; figure 2). GVs have been described solely

from lepidopteran hosts, whereas NPVs have been isolated from a wider range of insects.

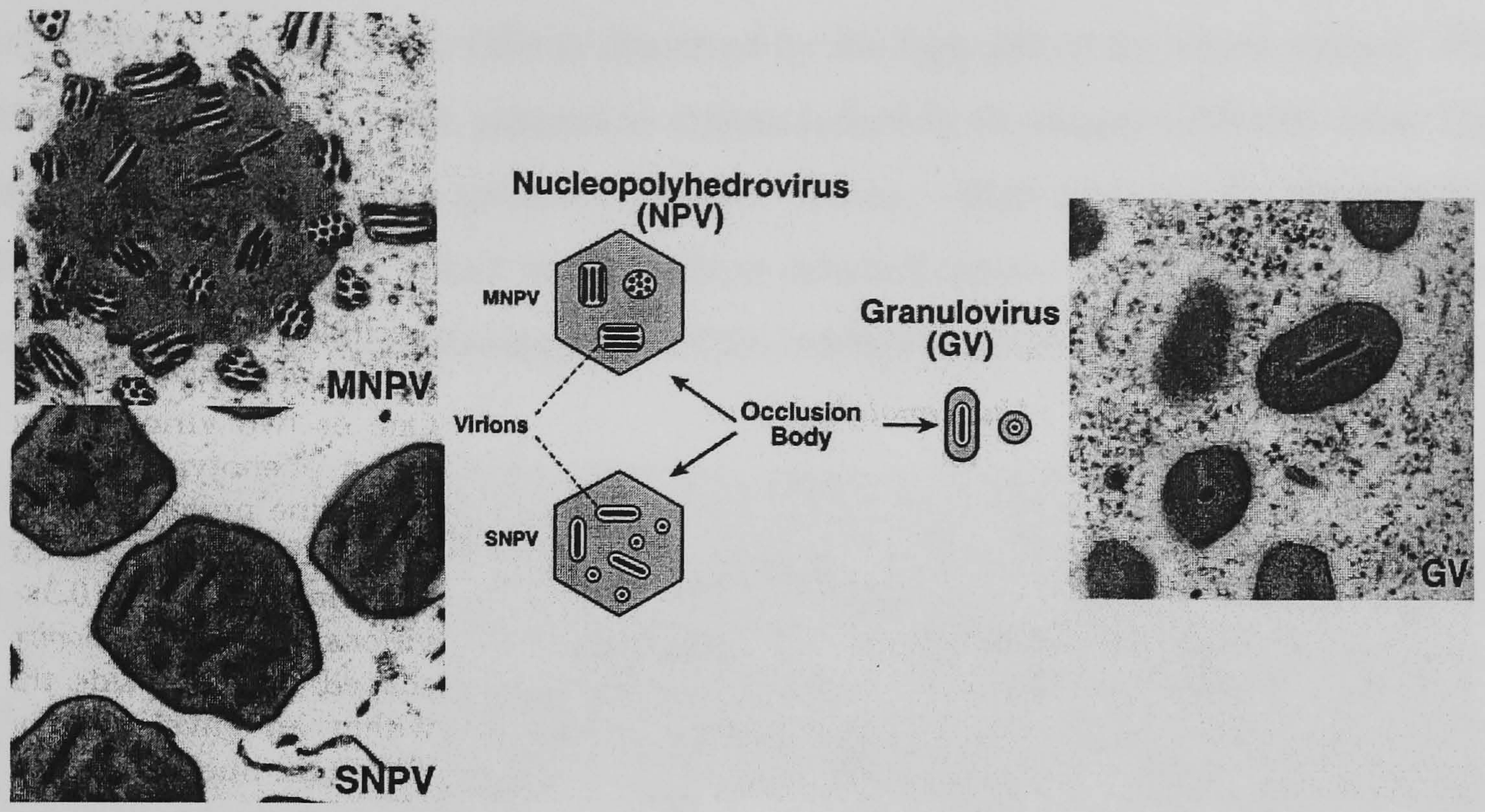


Figure 2.: Phenotypes of Baculovirus Occlusion Bodies. (Modified from Blissard *et al* 2000).

The taxonomical status of non-lepidopteran baculoviruses is still uncertain (Blissard et al., 2000). Although baculoviruses have been isolated from a large number of host species, only a fraction of them are listed in the VIIth ICTV report due to the lack of characterisation or the loss of original samples (Blissard et al., 2000). Prior to the Sixth ICTV report (Volkman et al., 1995), the *Baculoviridae* included a third group, the *Non Occluded Baculoviruses*. These viruses were grouped with the baculoviruses because of their rod-shaped, enveloped morphology. However, they do not form occlusion bodies and were therefore removed from the family (van Regenmortel et al., 2000). Examples of such viruses are the *Oryctes* rhinoceros beetle virus, which infects Coleoptera, and the gonad-specific virus and Hz-1 virus, both of which infect the moth *Helicoverpa zea* (Burand, 1999).

3. Life cycle - Pathogenesis

a) Host pathology

The baculovirus life cycle typically involves the production of two virus forms: the occluded viruses (OVs), which can persist in the environment and initiate host infection, and the budded viruses (BVs), which spread the infection between cells

within the host (Blissard & Rohrmann, 1990, Miller, 1997). Baculoviruses primarily infect larval stages (Federici, 1997). For lepidopteran hosts, caterpillars become infected when they ingest baculoviruses in the form of OBs (figure 3). After ingestion, the crystalline structure of the OBs is dissolved by the high pH of the insect midgut. This releases the virions, which proceed to initiate infection of midgut epithelial cells. This initial round of infection produces budded viruses, which disseminate the infection throughout susceptible insect tissues. These infected tissues eventually produce OBs, which are liberated upon disintegration of the caterpillar cadaver.

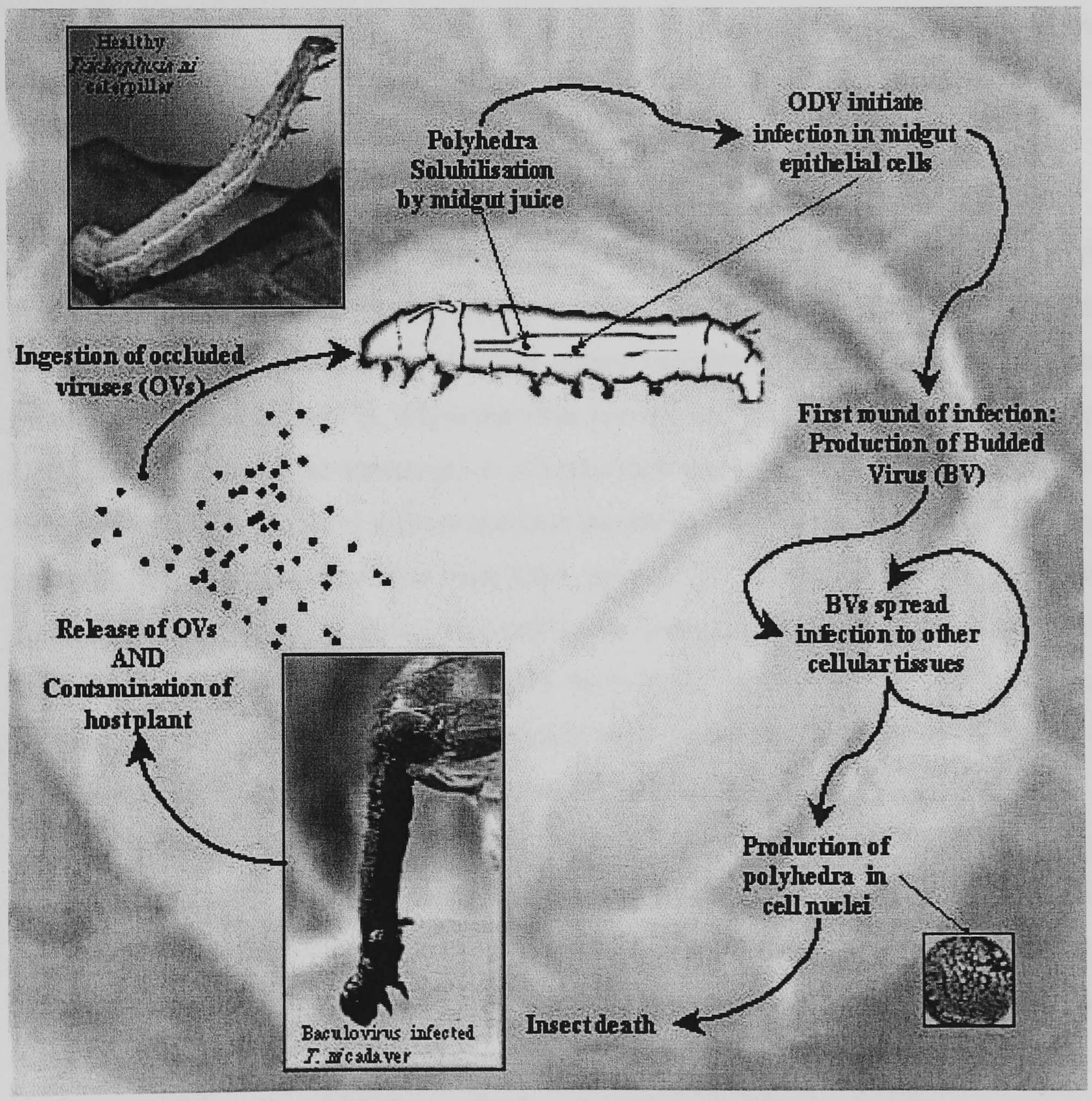


Figure 3.: Baculovirus Infection Cycle.

Baculoviruses are mainly transmitted horizontally by environmental contamination leading to naïve caterpillar ingesting occlusion bodies. It has to be noted that not all baculovirus infections cause the death of larval hosts; some infected hosts

survive to the adult stage and are able to reproduce (Kukan, 1999, Rothman & Myers, 1996). However, sublethal effects have been linked to non-lethal baculovirus infection as the surviving adults have generally lower reproductive success than non-infected adults (Matthews et al., 2002, Myers et al., 2000, Rothman & Myers, 1996). Furthermore, there are a number of examples where vertical transmission of baculoviruses has been demonstrated (Burden et al., 2002, Fuxa et al., 2002, Kukan, 1999). In these studies, adults challenged as larvae by baculoviruses were transmitting viruses to their offspring through the eggs. Not all the offspring bear the virus, or develop the disease. There is little knowledge of the exact mechanisms involved in virus latency or in the initiation of pathology, but it is thought that stress, such as overcrowding or food shortage, might increase susceptibility to the disease (Fuxa et al., 2002).

b) Viral replication at the cellular level

After the dissolution of the occlusion bodies, the occlusion derived viruses (ODVs) fuse with the membrane of the midgut cells (figure 4). The nucleocapsids (NCs) are directed towards the nucleus where baculovirus replication takes place (Williams & Faulkner, 1997). When the virus particle has reached the nucleus, the viral DNA is uncoated and the transcription of immediate early genes starts. At the onset of viral transcription, the virus utilises the host transcriptional machinery. Later on in the infection, baculoviruses use their own RNA polymerase encoded within their genome (Guarino et al., 1998b). Essential viral replication genes, such as DNA polymerase and helicase, have to be translated for DNA replication to take place. The new viral genomes are then package into nucleocapsids. And depending on the stage of infection, these can either mature as budded viruses or occluded viruses (figure 5).

Earlier in the infection the NCs are destined for the production of BVs. They exit the nucleus through transport vesicles. Once in the cytoplasm, the NCs are released from the vesicles (Williams & Faulkner, 1997). They progress to the cellular membrane from which they bud, thus acquiring a membrane as they leave the cells (figure 4). The budded viruses are the agents of secondary cellular infection. They spread the infection to all susceptible cells, which they actively enter using viral fusion proteins.

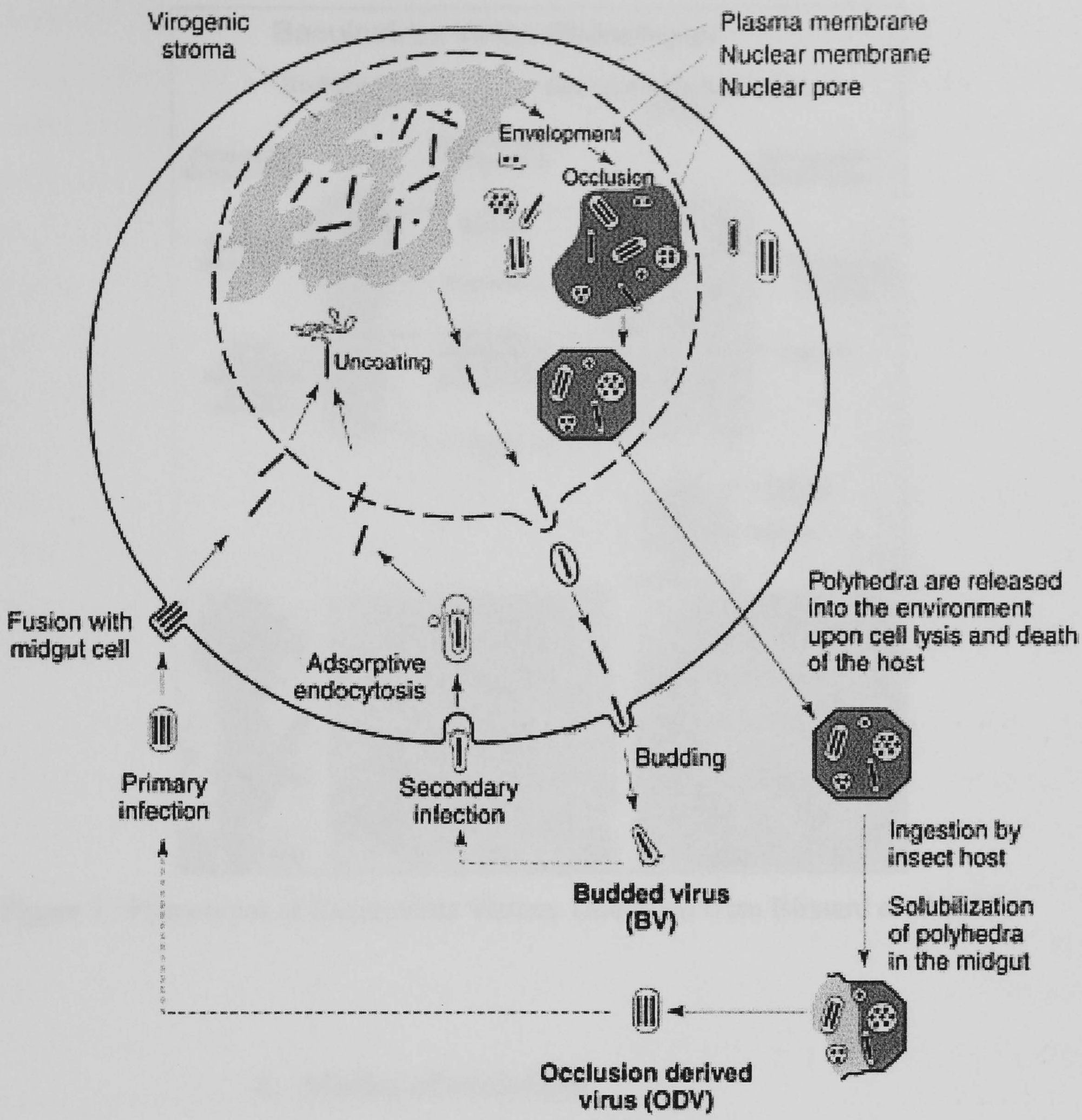


Figure 4: Baculovirus Replication at Cellular level (Encycolpedia of Virology, Rohrmann 1999, Nucleopolyhedrovirus)

At very late times in the infection the production of BVs is curtailed and replaced by the production of occluded viruses (Williams & Faulkner, 1997). The whole process occurs in the nuclei of infected cells (figure 4). The NCs acquire an envelope derived in part from the nuclear membrane and containing a suite of viral structural proteins such as the conserved protein p74 (Funk et al., 1997). The viral particles are then assembled along with the occlusion matrix protein polyhedrin (or granulin for the GVs) to form the crystalline structures of the OVs (figure 5). This process provokes a large swelling of the infected cell nuclei as the OVs accumulate. The nuclei eventually disintegrate and the occluded viruses are released from the cells (Williams & Faulkner, 1997).

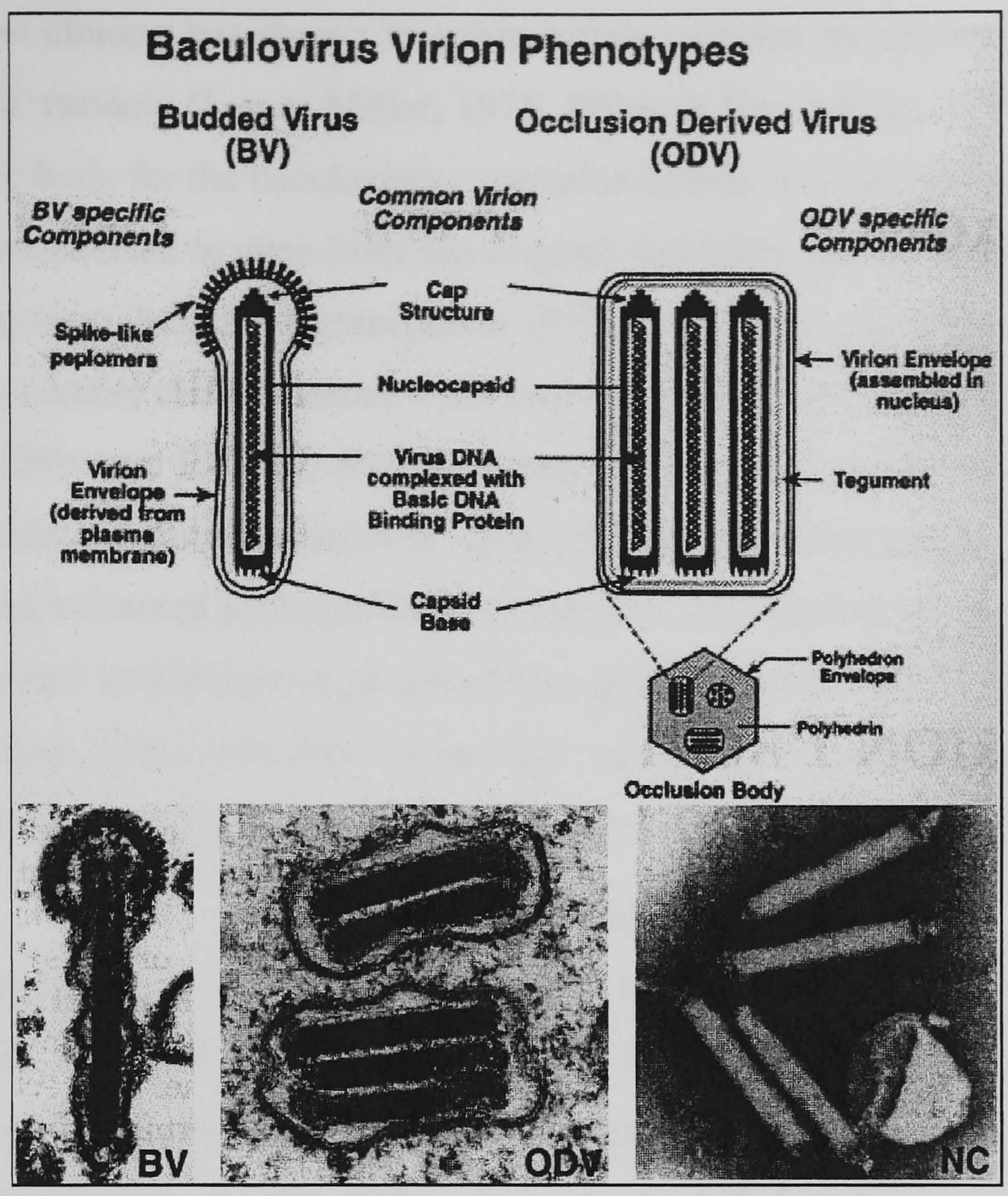


Figure 5.: Phenotypes of Baculovirus Virions. (Modified from Blissard *et al* 2000).

4. Modes of evolution

The evolution of baculoviruses is governed by the same rules as that of any organism. The results of evolution can be seen as the replacement genotypes. For evolution to happen, there must first be genetic variation. Variation can be created and maintained in different ways but has to be present before natural selection or genetic drift can act (Futuyama, 1998).

a) Evidence of genetic variation

We know very little about the mutation rate of baculovirus genomes during replication. However, it is reasonable to assume that, of the millions of particles are produced at the end of an infection, some will harbour new mutations. There is ample evidence of the existence of genetic variation in natural populations of baculoviruses. A single infected caterpillar of the pine beauty moth, *Panolis flammea*, was reported to contain 24 different genotypes (Hodgson et al., 2001). The original AcMNPV isolate was collected from a single *Autographa californica* larva over 20 years ago (Dr P.

Vail). *In vitro* cloning has shown that this isolate contains numerous genotypically distinguishable variants (Lee & Miller, 1978, Stiles & Himmerich, 1998), several of which form the basis for the baculovirus expression system used worldwide. In a recent study, 50 clones isolated *in vitro* from this original AcMNPV isolate were examined by restriction fragment length polymorphism (RFLP) analysis and tested in two host species. Seven distinct clones / strains were identified, 74% of the clones were of the L1 type variant, 12% were E2 type, 6% represented a previously unidentified variant and there were 3 unique genotypes (Stiles & Himmerich, 1998). Interestingly, several of the variants showed enhanced pathogenicity and speed of kill against two caterpillar pests, *Heliothis virescens* and *Helicoverpa zea* (Stiles & Himmerich, 1998). This wild-type AcMNPV isolate shows that there is enough variability in a natural population to respond to selection and adapt to different conditions, including a variety of hosts.

b) The origin of genetic variation

The natural creation of new genotypes can be mediated by different factors. *De novo* mutations can be incorporated due to DNA replication errors. Environmental factors such as UV light and chemicals can damage DNA. Baculovirus DNA is particularly susceptible to UV light when the occluded viruses are free in the environment. Indeed, the longer viruses are exposed to sunlight between infection cycles the less virulent they become (Ramoska et al., 1975). The need for higher doses of 'weathered' virus implies that accumulation of mutations inactivates a large portion of the natural virus population (Cory et al., 1997).

Recombination between nucleic sequences of viruses allows the shuffling of genotypes by re-assorting genes and alleles, similar to sex in cellular organisms. Exchanges of genetic material are known to occur among baculoviruses. Heterologous recombination can happen between genomes of different baculoviruses coinfecting the same cell. This happens most probably between viruses belonging to different strains of a viral population infecting a common host (Croizier & Ribeiro, 1992). But it can also happen between closely related viruses such as AcMNPV and BmNPV, which normally infect different host species but can be forced to coinfect the same cell (Kondo & Maeda, 1991). Combined with *de novo* mutations, recombination generates the variation on which natural selection acts.

Recombination can also mediate the acquisition of foreign genetic material. Illegitimate recombination between host genomes and baculoviruses has been observed in the form of diverse classes of transposable elements found in some baculovirus

genomes. The TED *gypsy* retrotransposon and the IFP2 transposon from *Trichoplusia ni* can spontaneously insert themselves in the genome of AcMNPV when this virus is grown in *T. ni* cells (Fraser et al., 1995, Friesen & Nissen, 1990). Similarly, lepidopteran Tc1-like transposable elements have been found to jump into the genome of CpGV (Jehle et al., 1995, Jehle et al., 1998). Furthermore, phylogenetic analysis of the envelope gene of *gypsy* retrotransposons showed that these *Metaviridae (Errantivirus)* have probably acquired this gene from baculoviruses, thus permitting viral infection (Malik et al., 2000, Rohrmann & Karplus, 2001).

Illegitimate recombination within the same genome is also responsible for sequence duplications. This leads to a number of genomic features, such as tandem repeats or gene families. The major class of baculovirus repeated sequences are the homologous regions (*hrs*), which are interspread along the genomes (Hayakawa et al., 2000) and contain themselves repeated motifs, the number of which has been observed to vary between different strains of the same virus (Chen et al., 2002). There are two main baculovirus gene families. The *iap* gene family is homologous to the eukaryotic antiapototic genes, some viruses have up to 4 gene copies (Ahrens et al., 1997, Hyink et al., 2002). The precise role of each of these copies remains to be determined. The second main gene family is the baculovirus repeated ORFs (*bro*). The copy number of *bro* genes varies a lot between viral genomes (Ayres et al., 1994, Kuzio et al., 1999) and also has been shown to differ even between viral strains (Lopez-Ferber et al., 2001).

Lastly homologous recombination is a natural mechanism in baculovirus infected cells, which is used extensively in genetic engineering of these viruses. The improvement of baculovirus insecticides by genetic modification has been the object of numerous research projects (e.g. Arana et al., 2001, Harrison & Boning, 2000, Maeda et al., 1991, Martens et al., 1995, McCutchen et al., 1991).

The introduction to a viral population of individuals with new genotypes, or migration, is also a way to increase genetic variation within a natural population. Occlusion bodies, which persist in the soil or on foliage, can be transported to other locations by animals (possibly including the host). The spread of baculovirus insecticides is also a means of importing new genotypes into a natural viral population (Munoz et al., 1997).

5. Phylogeny

Phylogenies are the means by which hierarchical relationships that have arisen through evolution can be represented. In systematics, the aim is to reconstruct trees that represent the evolution of species, by resolving the relationships between extant organisms and their ancestors. Morphology and molecular sequences can be used for this purpose. If these characters have evolved from the separation of species, the derived phylogenetic trees should reflect the species relationships. However, if this is not the case, the trees will only reflect the evolution of characters and not that of the species. This is why it is recommended that more than one set of characters is used, like different genes, and their congruence evaluated (Sanderson & Shaffer, 2002).

Viruses, unlike cellular organisms, often have extremely small genomes. This can limit the choice of suitable candidate genes for phylogenetic analyses. For example, in the retroviruses, the only regions available for family level phylogenetic studies are the protease and reverse transcriptase domains from the GAG gene (Herniou et al., 1998). In this case, the virus species phylogeny is the same as that of the genes, firstly because the small size of the genome makes it unlikely to be otherwise, and secondly because the phylogeny cannot be established based on any other set of independent characters.

However baculoviruses genomes are large and usually comprise more than a hundred genes. Thus there are numerous candidate genes for phylogenetic inference. This allows congruence between phylogenetic trees to be examined and should permit the construction of educated hypotheses about baculovirus species relationships. At the beginning of this project 6 complete baculovirus genome sequences were available in the databases. Since then the number has more than doubled and is still growing. It is therefore possible to investigate baculovirus evolution using whole genome datasets.

Traditionally, baculovirus phylogenies have been based on individual gene sequences. The polyhedrin/granulin (*polh*) gene, encoding the major matrix protein of the occlusion bodies, has been the most widely used (Bideshi et al., 2000, Bulach et al., 1999, Cowan et al., 1994, Liu & Maruniak, 1999, Zanotto et al., 1993) but other genes, such as DNA polymerase (*dpol*), *egt*, *gp41*, *chitinase*, *cathepsin*, *lef2* and *gp37* have also been used (Bulach et al., 1999, Chen et al., 1999, Chen et al., 1997, Clarke et al., 1996, Jin et al., 1999, Kang et al., 1998, Li et al., 2001, Liu & Maruniak, 1999). In general, these studies agree that the lepidopteran NPVs and GVs constitute distinct well-defined groups, and that the lepidopteran NPVs can be subdivided into two subgroups (figure 6), the group I NPVs including AcMNPV and OpMNPV, and the

group II NPVs typically including HaSNPV and SeMNPV (Bulach et al., 1999, Cowan et al., 1994, Zanotto et al., 1993).

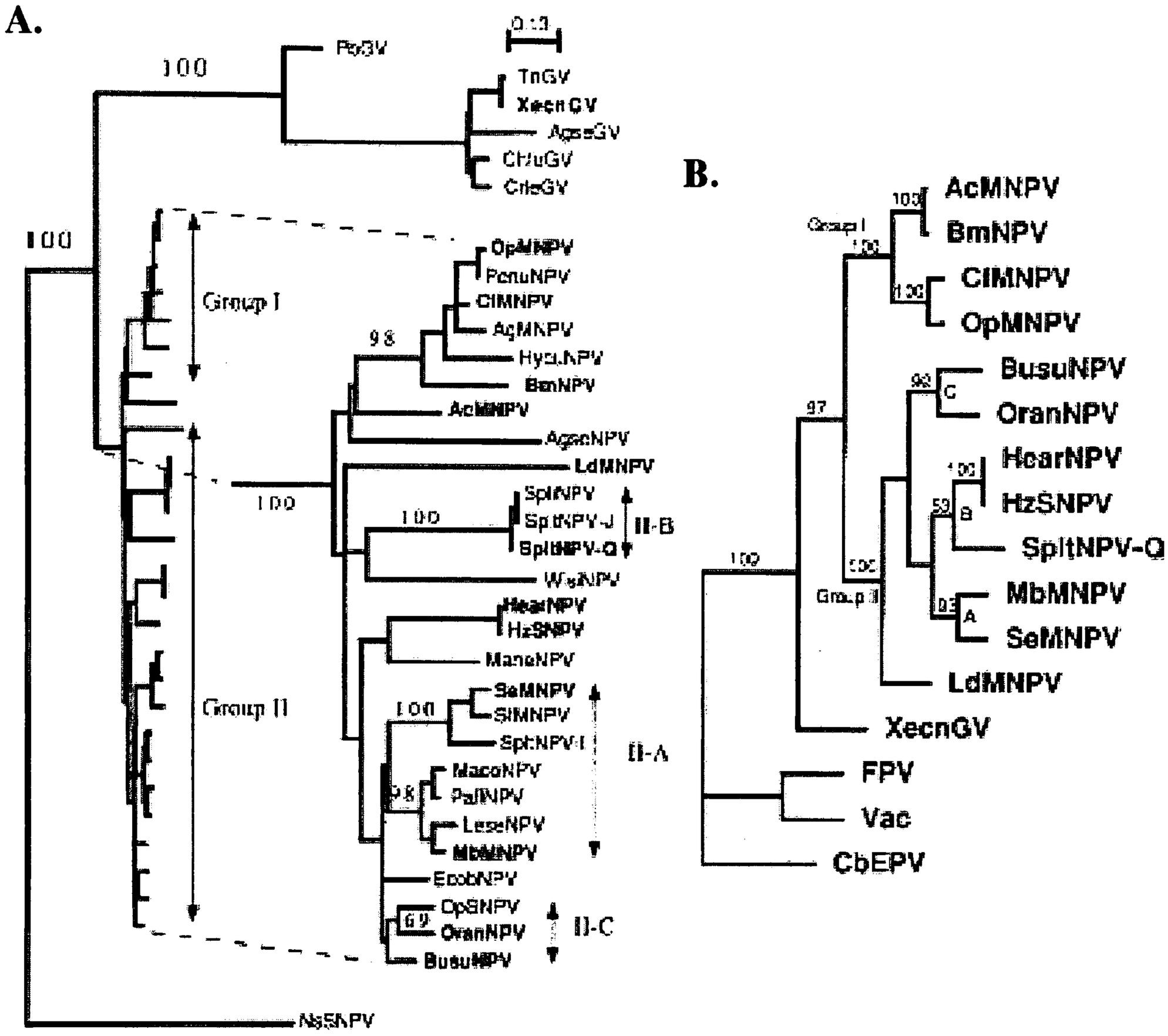


Figure 6: A. Baculovirus phylogeny based on polyhedrin/granulin amino acid sequences. B. Baculovirs phylogeny based on DNA-polymerase amino acid sequences. (Bulach et al 1999).

Almost all phylogenetic studies have been based on sequences from lepidopteran baculoviruses. Little work has been done to try to investigate the position of non-lepidopteran baculoviruses. Resolving the relationships between viruses isolated from Hymenoptera, Diptera, Crustacea and Lepidoptera would greatly enhance our understanding of the evolution of the virus family *Baculoviridae*. Early amino acid sequencing of the polyhedrin protein of *Neodiprion sertifer nucleopolyhedrovirus* (NeseNPV; Rohrmann et al., 1981) and more recent sequencing of the complete DNA sequence of the gene (Zanotto et al., 1993), showed that the *polh* sequence of this hymenopteran virus is quite divergent from that of the lepidopteran viruses, including

NPVs and GVs. These phylogenies based on the occlusion body protein imply that the hymenopteran virus is from an ancient lineage. However, these trees should be treated with caution until further evidence is available to support this result, as the *polh* gene could be prone to selective pressure and adaptation to the host due to its pivotal role in baculovirus occlusion. More recently, phylogenetic analyses based on the p74 gene that include sequences from the dipteran virus *Culex nigripalpus nucleopolyhedrovirus* (CuniNPV) also showed that this virus is very divergent from the lepidopteran viruses (Moser et al., 2001). There are currently no baculovirus phylogenies including viruses from the 3 arthropods orders, to understand the earlier evolutionary history of the group.

B. Thesis Outline

The thesis has been organised into three chapters, beginning in chapter 1 with a general introduction on baculoviruses. My studies of baculovirus evolution took two directions as illustrated by chapters 2 and 3.

1. Chapter 2: The use of complete genomes to investigate baculovirus evolution

Chapter 2 provides detailed phylogenetic analyses based on complete genome sequences to understand baculovirus evolution. It is based on two publications. The first paper described the use of gene sequences, gene order and gene content data set to understand the relationships between nine completely sequenced baculoviruses (Herniou et al., 2001). The second paper was an update and review of the methods using 13 genomes (Herniou et al., 2003). The chapter itself is a fusion and revision of both papers. It is divided into four sections. Section A gives an account of the genes forming the core of baculovirus genomes. Section B presents the methods of phylogenetic analyses based on gene sequences, gene order and gene content, and the results of the analyses based on nine and thirteen genomes. Section C relates the integration of the baculovirus phylogeny tree and comparative genomics methods to understand the evolution of genome content and discover putative functional linkage between genes.

2. Chapter 3: Sampling Baculovirus Diversity and Evolution

Chapter 3 aims to explore the diversity of the *Baculoviridae* by acquiring molecular sequences from as many isolates as possible in order to reconstruct their evolutionary relationships. The chapter is divided into six sections. Section A relates the different approaches, which have already been taken to understand baculovirus diversity including the characterisation of viral isolates, baculovirus nomenclature and phylogenetic studies. Section B presents the materials and methods utilised to acquire and analyse new baculovirus sequences from a variety of isolates. Section C focuses on the relationships between baculoviruses from different insect orders. Section D

investigates the relationships between baculoviruses isolated from lepidopteran hosts with an application of the phylogenetic species concept to cluster isolates into species. Section E examines the evolution of host use in lepidopteran baculovirus. Section F summarise the advances obtained by the utilisation of evolutionary approaches on complete genomes to understand baculovirus biology.

II. Chapter Two: The Use Of Complete Genomes To Investigate Baculovirus Evolution

A. Baculovirus Genomes

Baculoviruses have circular double stranded DNA genomes (figure 7). Their sizes range between 179 kbp for XcGV and 101 kbp for PxGV (Hashimoto et al., 2000, Hayakawa et al., 1999). It therefore represents a non-trivial effort to determine and annotate the complete sequences of baculovirus genomes.

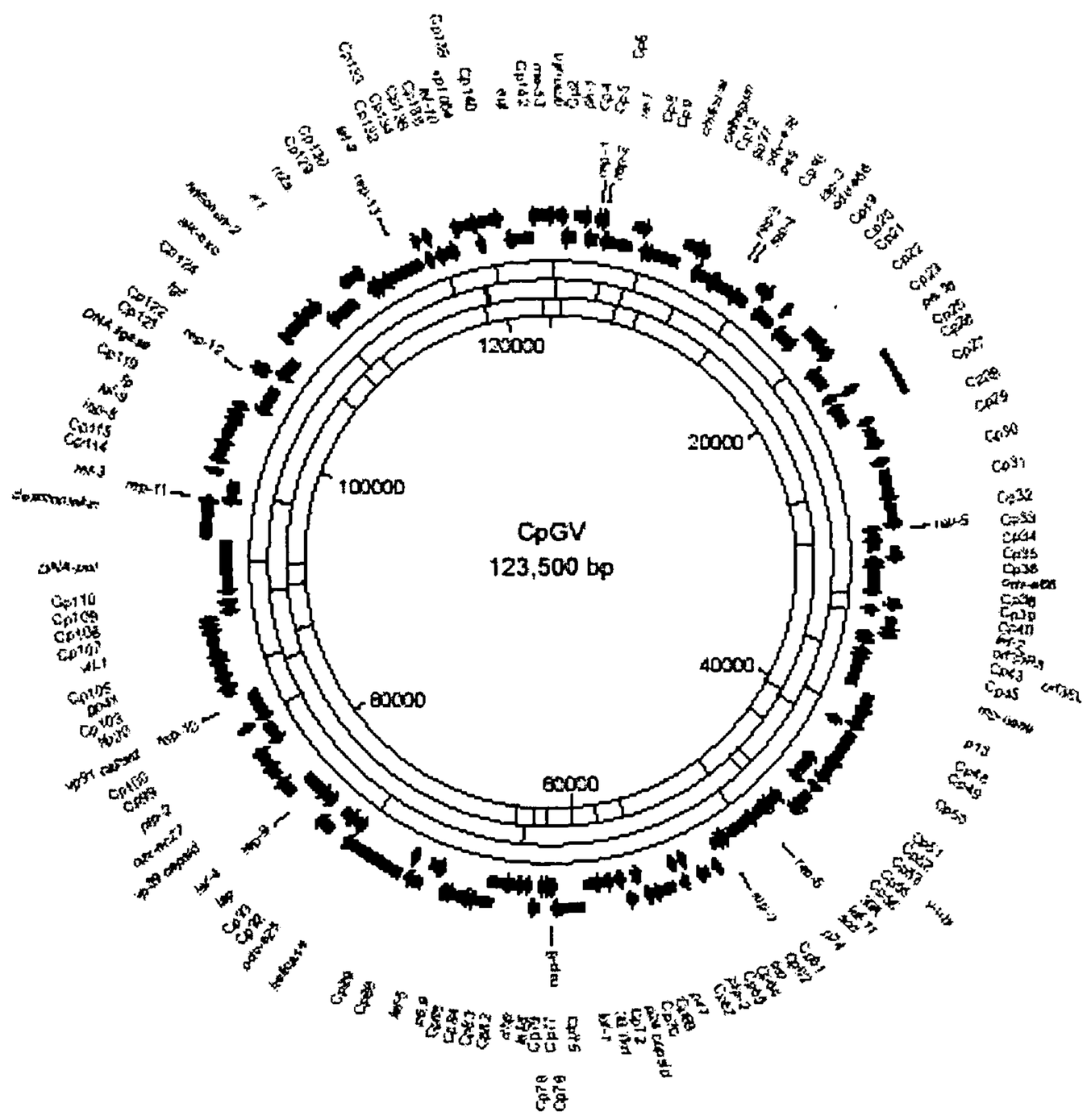


Figure 7.: Circular Map of the CpGV Genome (Luque *et al*, 2001)

The first baculovirus to be completely sequenced was *Autographa californica* NPV (AcMNPV; Ayres et al., 1994). The number of complete sequences has since grown rapidly, reaching 13 genomes at the time of analyses, including three GVs (Hashimoto et al., 2000, Hayakawa et al., 1999, Luque et al., 2001), nine lepidopteran NPVs (Ahrens et al., 1997, Ayres et al., 1994, Chen et al., 2001, Chen et al., 2002, Gomi et al., 1999, Hyink et al., 2002, IJkel et al., 1999, Kuzio et al., 1999, Pang et al., 2001) and one dipteran NPV (Afonso et al., 2001). Since the time of the analyses, four more lepidopteran baculovirus genomes have been made available: two NPVs from *Mamestra configurata* (Li et al., 2002a, Li et al., 2002b), another NPV from *Rachiplusia ou* (AY145471, Bonning & Harrison, unpublished) and one GV from *Phthorimaea operculella* (AF499596, Croizier, L. *et al,* unpublished). Each complete

sequence has provided a wealth of data about the complement of genes present in each individual virus. However, whole genome sequences, when compared between viruses can also suggest how virus family members are related to one another, and how they have evolved since their last common ancestor. This chapter will assess the insights into baculovirus biology and evolution that can be gained by combining comparative genomic and phylogenetic approaches.

1. Core Baculovirus Genes

The 13 completely sequenced baculoviruses have 30 identified genes in common. A putative function can be assigned to 20 of these (table 1). They include genes that act at various stages of the baculovirus infection cycle, and can be grouped into five functional categories: RNA transcription, DNA replication, structural proteins, auxiliary proteins, and proteins of unknown function.

Table 1: Genes present in all baculovirus genomes

Gene function	**Genes present in all baculoviruses**	**Additional genes present in all lepidopteran baculoviruses**
Replication	*lef-1 (ac14)*[†]*, lef-2 (ac6), dnapol (ac65), helicase (ac95)*	*dbp1 (ac25), lef-3 (ac67), ie-1 (ac147), me53 (ac139)*
Transcription	*p47 (ac40), lef-8 (ac50), lef-9 (ac62), vlf-1 (ac77), lef-4 (ac90), lef-5 (ac99)*	*pp31/39K (ac36), lef-6 (ac28), lef-11 (ac37)*
Structural Proteins	*bfp (ac23, ld130), gp41 (ac80), odv-ec27 (ac144), odv-e56 (ac148), p6.9 (ac100), p74 (ac138), vp91/p95 (ac83), vp39 (ac89), vp1054 (ac54)*	*fp25K (ac61), odv-e18 (ac143), odv-e25 (ac94), odv-e66 (ac46), pk1 (ac10), polh (ac8), ac82(bm68)*
Auxiliary *	*alk-exo (ac133)*	*fgf (ac32), ubiquitin (ac35)*
Unknown	*38K (ac98), ac22, ac68, ac81, ac92, ac96, ac109, ac115, ac119, ac142*	*38.7K (ac13), ac29, ac38, ac53, ac66, ac75, ac76, ac78, ac93, ac106, ac110, ac145, ac146, p40 (ac101), p12 (ac102), p45 (ac103)*

* Auxiliary genes are genes that are not directly involved in viral gene expression, genome replication, or in the formation of progeny virus particles (O'Reilly, 1997). † ORF number of the AcMNPV homologue.

a) Transcription

Baculovirus gene expression is temporally regulated primarily at the level of transcription (Friesen, 1997). Baculovirus transcription is typically categorised into

three classes: early, late, and very late. Host RNA polymerase II transcribes early genes (Friesen, 1997), but a viral RNA polymerase transcribes late and very late genes (Fuchs et al., 1983, Guarino et al., 1998a). The AcMNPV RNA polymerase activity is composed of four subunits, all of them viral proteins (Guarino et al., 1998a). The genes, which encode these, *lef-4, lef-8, lef-9* and *p47*, are present in all sequenced baculoviruses. *Lef-8* and *lef-9* (Lu & Miller, 1994) display homology to subunits of the DNA-dependent RNA polymerases of prokaryotes and eukaryotes, with *lef-8* containing a motif conserved in β and β′ subunits of RNA polymerises, which is thought to form part of the catalytic site of the enzyme (Passarelli et al., 1994). The lef-4 protein of AcMNPV, which has guanylyltransferase (Guarino et al., 1998a), RNA 5′-triphosphatase, and ATPase (Jin et al., 1998) activities, is proposed to provide a capping function for RNA transcribed by the baculovirus polymerase (Jin et al., 1998). The specific function of *p47* remains to be determined.

Two other genes implicated in transcription, *lef-5* and *vlf-1,* are in the core gene set. The function of *lef-5* is not clear. V*lf-1* is required for very late gene transcription in AcMNPV and has sequence homology to a gene family of integrase/resolvases (McLachlin & Miller, 1994). Studies indicate that the protein vlf-1 physically interacts with promoter sequences, allowing for the 'burst' of expression observed for very late promoter-driven genes (Yang & Miller, 1999).

b) DNA Replication

Baculoviruses DNA replication takes place in the infected cell nucleus where all the cellular machinery is available. However, the viruses do not seem to utilise the cell DNA replication system, as they carry their own complement of genes encoding DNA replication proteins. Four of these genes, *DNA polymerase* (*dpol*), *DNA helicase* (*p143, hlx*), *lef-1* and *lef-2*, are found in all sequenced genomes and are essential for the replication of baculoviruses (Kool et al., 1994, Liu & Carstens, 1999, Lu & Miller, 1995).

The baculovirus *dpol* gene is homologous to the DNA polymerase *alpha* of eukaryotes, the *polII* gene of archaea and DNA polymerase genes from other DNA viruses such as herpesviruses, poxviruses and phicodnaviruses (Villareal & DeFilippis, 2000). Its activity is to replicate the DNA of baculoviruses (Tomalski et al., 1988). The baculovirus *hlx* gene has low homology to helicase genes of other organisms. Its activity is to unwind baculovirus DNA at the replication fork (McDougal & Guarino, 2000). *Lef-1* encodes a DNA primase (Barrett et al., 1996, Mikhailov & Rohrmann,

2002). The Lef-1 and Lef-2 proteins are thought to form hetero-oligomeric complexes during baculovirus replication (Evans et al., 1997).

These 4 genes assure the basic activities required for DNA replication. The helicase unwinds the DNA and creates the fork. The primase initiates the start of replication by the DNA polymerase on the lagging strand by synthesizing a RNA primer. In cellular systems, the last protein required for DNA replication is the single stranded DNA binding protein (SSB; Lewin, 2000). The *lef-3* gene has been shown to encode a SSB, which stabilises the single stranded DNA during the replication (Hang et al., 1995). The mosquito virus CuniNPV seems to lack a homologue of this gene (table 1).

c) Structural Proteins

Newly replicated baculovirus genomes are packaged with viral proteins to form nucleocapsids (NCs). Four genes implicated in NC structure or assembly are conserved amongst the 13 baculovirus genomes: *p6.9, vp39, vp1054*, and v*p91* (table 1; Funk et al., 1997). The p6.9 protein is a small basic protein, which is thought to facilitate condensation of the DNA associated with packaging within the nucleocapsid (Tweeten et al., 1980, Wilson et al., 1987). Vp39 is the major capsid protein (Blissard et al., 1989, Pearson et al., 1988, Thiem & Miller, 1989). Vp1054 and Vp91 are both associated with budded virus and occluded virus (Olszewski & Miller, 1997a, Russell & Rohrmann, 1997).

The baculovirus F protein (*bfp*, *ld130*) (Pearson & Rohrmann, 2002, Westenberg et al., 2002) is an envelope fusion protein found in infected-cell membranes and in BV envelopes (IJkel et al., 2000, Pearson et al., 2000). It is required for fusion of the budded virus envelope with the host endosomal membranes, to release nucleocapsids for initiation of infection. Interestingly, for viruses such as AcMNPV, BmNPV and OpMNPV this fusion activity is provided by an unrelated BV envelope protein, gp64, even though they possess a *bfp* gene (Blissard & Rohrmann, 1989, Blissard & Wenz, 1992, Whitford et al., 1989). Gp64 homologues are more closely related to one another than the less-conserved *bfp* homologues (Rohrmann & Karplus, 2001). It has therefore been suggested that *bfp* represents an ancestral envelope fusion protein gene, whose activity has been replaced in the group I NPVs by *gp64* (Pearson et al., 2001), a process known as non-orthologous gene displacement (Galperin & Koonin, 2000). Presumably, *bfp* homologues in *gp64*-containing genomes have evolved some other essential viral function (Pearson et al., 2001).

Odv-ec27, Odv-ec56, p74 and Gp41 are all conserved proteins found in association with occluded viruses (Braunagel et al., 1996a, Braunagel et al., 1996b, Kuzio et al., 1989, Whitford & Faulkner, 1992). Odv-ec27 has homology to cellular cyclins and interacts with cyclin-dependent kinases and other cell cycle-related host proteins in *in vitro* assays, leading to the suggestion that this protein is a multi-functional viral cyclin (Belyavskyi et al., 1998). P74 is important for the oral infectivity of occluded viruses (Faulkner et al., 1997), while the Gp41 glycoprotein is required for egress of nucleocapsids out of the infected-cell nucleus to form budded viruses (Olszewski & Miller, 1997b).

Although CuniNPV is an occluded virus, it doesn't seem to harbour a homologue of the *polyhedrin/granulin* gene. The gene encoding CuniNPV major occlusion body protein does not bear any sequence similarity to the *polh* gene sequences (Moser et al., 2001).

d) Auxiliary Proteins / Unknown Function

The only auxiliary protein conserved amongst all 13 sequenced baculoviruses is a putative alkaline exonuclease (alk-exo). In addition, another 10 genes of unknown function are conserved among all 13 genomes. The broad conservation of these genes strongly suggests they provide essential functions for baculovirus replication. They are prime candidates for future molecular and biochemical studies. Preliminary data on homologues of *ac109* (Hu et al, SIP meeting 2002) and *ac119* (Kikhno et al., 2002) showed that they both encode proteins of the OV, but their precise role in baculovirus infection remains unclear.

2. Lepidopteran vs Dipteran Baculoviruses.

It is worth noting that a much greater number of genes are conserved among all the lepidopteran baculovirus genomes. An additional 32 genes are found in all lepidopteran viruses sequenced to date that are apparently absent from the mosquito CuniNPV genome (table 1). Some of these differences may reflect a different spectrum of genes required to infect a lepidopteran compared to a dipteran host. Alternatively, they may reflect the very different biology of the CuniNPV compared to the lepidopteran viruses that have been sequenced. Unlike the lepidopteran baculoviruses, CuniNPV is restricted to replication in a subset of midgut cells (Moser et al., 2001); therefore, these 'additional' lepidopteran genes may be involved in utilisation of various host tissues. However, it is important to bear in mind that the biology of these viruses

might not be as distinct as the sequence data appear to indicate. Phylogenetically unrelated genes may provide similar functions in different virus groups. One example of this, the *bfp* / *gp64* genes, has been discussed already. The baculovirus genes that prevent the programmed death of infected cells are another example. Two unrelated families of anti-apoptotic proteins are found in baculoviruses: p35 homologues and IAPs (Clem, 1997). All lepidopteran baculoviruses have at least one *iap* gene, while AcMNPV, BmNPV, SpltNPV, and CuniNPV all have functional or putative *p35* genes. Finally, it is possible that homologues of some of the lepidopteran virus core genes are present in CuniNPV, but too highly diverged to identify by BLAST searches (Afonso et al., 2001).

The minimal-gene-set concept has been developed to refer to the minimal group of gene necessary to replicate a cell. Comparative analyses of bacterial genomes showed that most of the genes included in the set were orthologues, but that a number were unrelated genes providing similar functions. This called for a widening of the concept to the more inclusive notion of minimal set of functions (Koonin, 2000).

Here an attempt has been made to identify the essential core set of genes that defines baculoviruses. However the 30 genes that have homologues in all the 13 genomes might under-represent the number of essential functions necessary to reproduce a baculovirus. The lack of *lef-3* in CuniNPV for example raises the question of whether another CuniNPV gene provides the SSB function or whether this virus uses the enzyme coded by the host. Application of the minimal-gene-set concept to the baculoviruses has so far led to 30 orthologous genes present in all the genomes. The function of a number of these genes remains to be elucidated. However they are surely essential to viral replication. The fact that baculoviruses carry around 150 genes in their genomes seems to suggest that the minimum set of functions necessary for viral replication might be higher than 30 genes.

B. Use Of Whole Genome Sequence Data To Infer Baculovirus Phylogeny

To circumvent the problem of conflicting single gene trees, entire genomes can be used to reconstruct baculovirus phylogenies. Several different approaches have been explored to take advantage of complete genome sequences. Genomes contain multiple levels of phylogenetic information. In addition to the nucleotide and amino acid sequences, complete genomes also contain structural information such as the order of genes or gene composition (Koonin et al., 2000). The rationale for using such characters for phylogenetic inference is that two genomes sharing similar organisation are more likely to have inherited it from a common ancestor than by chance. An advantage of evaluating evolution on a genomic scale is that the phylogenetic inferences are then based on the complete genetic make-up of each species.

The inference of phylogeny based on either nucleotide or amino acid sequences is greatly extended by the ability to use all genes in common between the genomes of interest. There are two potential approaches: to analyse each gene separately and derive a consensus from the resulting phylogenies, or to concatenate the genes and analyse them together. In principle, the second approach should give better results because each gene contributes to the overall phylogenetic signal and a synergistic effect is produced by combination of all the signals. This approach has proved useful for the phylogenetic analysis of metazoans based on mitochondrial genomes (Nikaido et al., 2001) and for the phylogeny of herpesviruses (Albà et al., 2001, Mcgeoch et al., 2000). It has also been successful when employed on a larger scale for the analysis of prokaryotic genomes (Sicheritz-Pontén & Andersson, 2001).

Gene order was first used in phylogenetics when the sequences of some mitochondrial genomes became available (Boore et al., 1995). Since then several approaches have been developed to reconstruct phylogenies based on gene order, using either distance methods or parsimony (Blanchette et al., 1999, Boore & Brown, 1998). Mitochondrial genomes however constitute a serious challenge because their small size limits the number of possible rearrangements (Le et al., 2000). The availability of sequences from larger, closely related genomes, such as herpesviruses and plant chloroplasts has provided new opportunities for the application of such approaches (Bourque & Pevzner, 2002, Hannenhalli et al., 1995).

Evolution of genome content is one of biology's most fundamental questions. How genes are inherited, acquired, lost, exchanged, duplicated and evolve are questions that have become predominant in recent years (Eisen, 2002, Kreitman & Gaasterland, 2001). It is not surprising therefore that the completion of recent genome projects has provoked renewed interest in these questions, and phylogenetic analyses based upon the presence or absence of genes, gene families or protein domains have been carried out on a range of microorganisms (Albà et al., 2001, Fitz-Gibbon & House, 1999, Lin & Gerstein, 2000, Montague & Hutchison, 2000, Snel et al., 1999, Snel et al., 2002, Tekaia et al., 1999)

1. Methods Of Phylogenetic Reconstruction Based On Complete Genomes

a) Baculovirus Sequences

Phylogenetic methodologies were developed and evaluated to reconstruct the evolutionary relationship of baculoviruses based on their complete genome sequences. In the first instance the analyses were performed on 9 complete genomes (table 2). These were all the genomes available at the time of the project development. This permitted the evaluation of the feasibility of the analyses.

The methods were tested again when 4 more genomes became available (table 2). Redoing the complete genome analyses using 13 genomes, allowed testing of the robustness of the methodologies, when more taxa are added to the data sets. Furthermore this second analysis includes the baculovirus isolated from the mosquito *Culex nigripalpus*, which is extremely divergent to the other baculovirus (Moser et al., 2001). Including this virus permitted to test how the different methods behave when a divergent taxon is incorporated to data sets.

Table 2: Complete baculovirus genome sequences

Baculovirus	Abbreviation	Size (bp)	no of genes	Accession	Reference
Autographa californica nucleopolyhedrovirus	AcMNPV	133894	155 (1)*	L22858	Ayres *et al,* 1994
Bombyx mori nucleopolyhedrovirus	BmNPV	128413	140 (5)	L33180	Gomi *et al,* 1999
Orgyia pseudotsugata nucleopolyhedrovirus	OpMNPV	131990	153 (1)	U75930	Arhens *et al*, 1997
Lymantria dispar nucleopolyhedrovirus	LdMNPV	161046	164 (1)	AF081810	Kuzio *et al*, 1999
Spodoptera exigua nucleopolyhedrovirus	SeMNPV	135611	140 (1)	AF169823	Ijkel *et al*, 1999
Helicoverpa armigera nucleopolyhedrovirus	HaSNPV	131403	135 (1)	AF271059	Chen *et al*, 2001
Xestia c-nigrum granulovirus	XcGV	178733	181	AF162221	Hayakawa *et al*, 1999
Plutella xylostella granulovirus	PxGV	100999	120	AF270937	Hashimoto *et al*, 2000
Cydia pomonella granulovirus	CpGV	123500	143	U53466	Luque *et al*, 2001
Epyphias posvittana nucleopolyhedrovirus †	EppoMNPV	118584	136 (1)	AY043265	Hyink *et al*, 2002
Helicoverpa zea nucleopolyhedrovirus †	HzSNPV	130869	139	AF334030	Chen *et al*, 2002
Spodoptera litura nucleopolyhedrovirus †	SpltNPV	139342	141 (3)	AF325155	Pang *et al*, 2001
Culex nigripalpus nucleopolyhedrovirus †	CuniNPV	108252	109	AF403738	Afonso *et al*, 2001

*The number in parenthesis indicates genes not recognised when the original sequence paper was published, or correspond to fused genes. †Viral genomes added for the second round of phylogenetic analyses

b) Gene Homology Relationships

A key step in understanding relationships between genomes is to determine which gene homologues represent orthologues (descended from the same gene copy in the last common ancestor of 2 species) or paralogues (descended from different genes in the last common ancestor, due for example to gene duplication). For a gene phylogeny to reflect the species phylogeny, orthologues must be used (Thornton & DeSalle, 2000). Evaluating gene orthology is therefore crucially important, as this determines which genes are used for all subsequent studies.

The genes identity, homology and position within each genome were taken from the publications describing the sequence annotations (appendix 1). The *bro* (baculovirus repeated ORF) gene family was deliberately omitted from the analyses because it is highly repeated in some baculovirus genomes, making it problematical to establish

orthology relationships (i.e. it is very difficult to judge which *bro* gene is the orthologue of which).

The different analysis methods use different subsets of baculovirus genes (figure 8). The data set for phylogenies based on gene sequences (nucleotides or amino acids) is the set of genes common to all genomes. The 9 genomes included in the first round of phylogenetic analyses have 63 genes in common, whereas the 13 genomes used for the second set of analyses only share 30 genes. Neighbour pair analysis is based on the order of the very same set of genes as the gene-sequence phylogenies. Breakpoint analysis is also based on gene order, but considers all genes shared by at least 2 genomes. Gene content analysis is based on all genes except those that are present in all genomes. Thus it is based on a non-overlapping gene set to that used for the gene sequence analyses.

c) Phylogenetic Inference Based On Gene Sequences

For each gene in common to the set of genomes (63 for the analyses based on 9 genomes and 30 for the analyses based on 13 genomes), amino acid sequences were aligned with ClustalW (Thompson et al., 1994) using default parameters and the Blosum matrix. The alignments were checked and refined by eye using MacClade 4 (Maddison & Maddison, 2000) prior to being compiled in a single Nexus file. Each gene represents a defined subset of this file.

Maximum parsimony analyses were performed in PAUP * (Swofford, 2001) . Gaps were treated as missing data. Phylogenies were estimated using a PAM-weighted amino acid step matrix (Telford, 2000). For the analyses based on 9 genomes, it was possible to perform exhaustive searches to find the most parsimonious trees. In the case of the 13 taxa analyses this was no longer possible. Heuristic searches with 10 random additions were therefore performed instead.

Phylogenetic trees were reconstructed either by analysing each gene separately or by concatenating the alignments and performing one analysis. By first analysing each gene separately, the heterogeneity of the phylogenetic signal contained within each gene can be evaluated. If all the individual genes give the same phylogenetic tree, then the signal contained in each gene is as good as that of all of them combined. However, if each gene gives a different tree, suggesting weak individual data sets, it is likely that combining data sets would reinforce the signal. Strong disagreements between individual gene phylogenies could indicate that those genes have had different evolutionary histories.

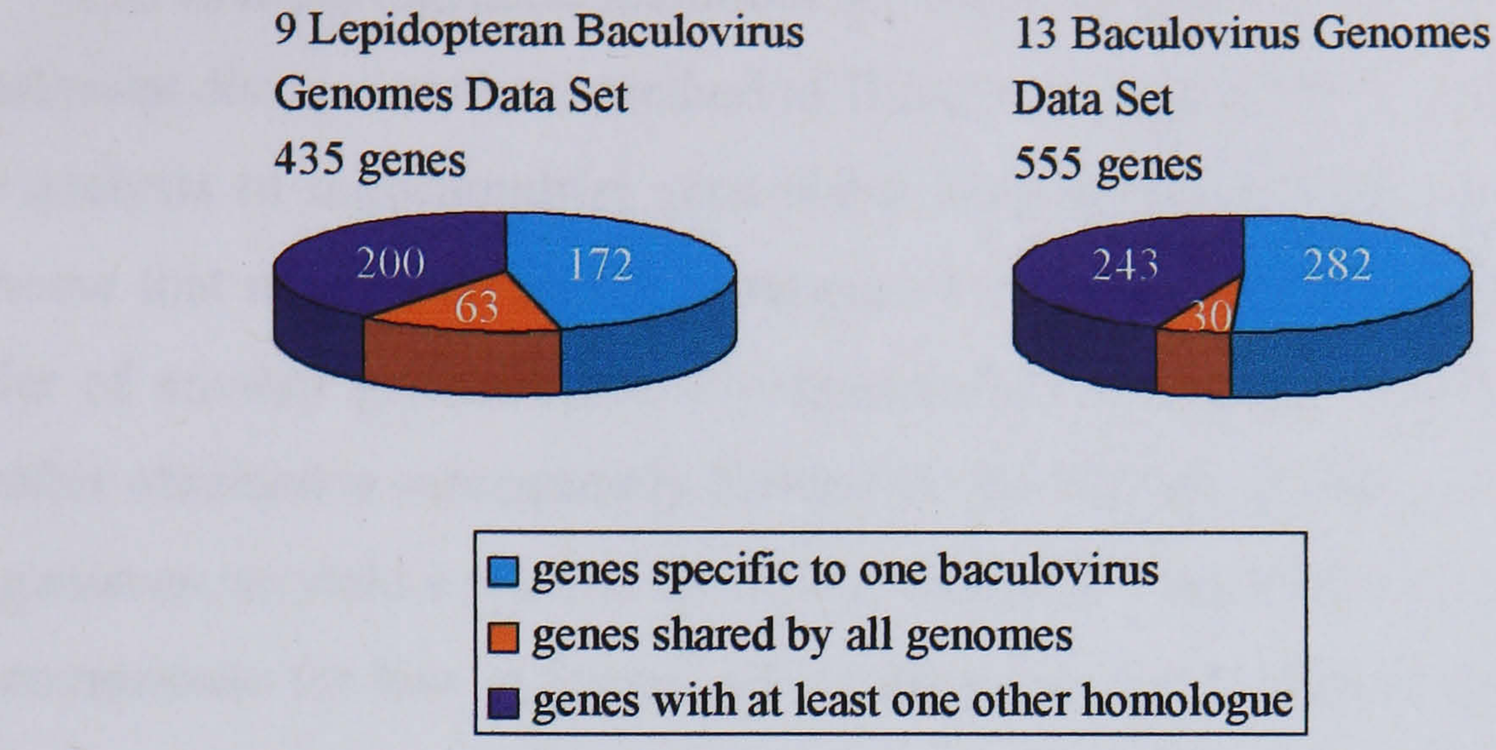
In the case of the single gene analyses, the most parsimonious trees (MPTs) were also retained to calculate majority rule consensus trees, which summarise the results of the analyses and can be compared to the results obtained with the combined genes. Some single gene analyses gave more than 1 MPT. The majority consensus tree was calculated taking into account such cases, by giving a weight of 1 tree per single gene analysis. In the event of a phylogenetic analysis giving 2 trees, each would be weighted 1/2 in the calculation of the majority rule consensus tree. Therefore the majority rule consensus tree is the summary the single gene analyses and the numbers above the branches represent the percentage of genes that gave rise to each branch.

d) Phylogenetic Inference Based On Gene Order

Gene parity plots, which compare the positions of orthologous genes on different genomes, have been widely used to assess synteny conservation between baculovirus genomes (Chen et al., 2001, Hashimoto et al., 2000, Hu et al., 1998, IJkel et al., 1999, Li et al., 2002a, Li et al., 2002b, Luque et al., 2001). They have shown that the gene order of GVs is fairly well conserved, but quite different to that of the NPVs (Hashimoto et al., 2000, Luque et al., 2001), that the group I NPVs also share very similar gene arrangements, and that gene order is far less conserved amongst the group II NPVs (Chen et al., 2001, Hu et al., 1998, IJkel et al., 1999). Gene parity plots demonstrate the utility of synteny comparison for assessing the relationships of baculoviruses. Although the recent developments of identity gene parity plots, that combine in the same graph the gene syntenies and identities, give a better idea of genomic distances (Li et al., 2002a), the gene parity plots only provide a qualitative estimate and their use is limited for phylogenetic studies due to lack of quantitative data. Other methods are necessary to interpret gene order within an evolutionary context.

Two approaches have been used to infer phylogenies based on gene order. The first, relative breakpoint analysis, is based on a comparison of gene order between pairs of genomes, and is therefore based on all genes present in more than one virus, whereas the second, neighbour pair analysis, is based on the arrangements of the genes common to all baculovirus genomes (figure 8).

A. Homology distribution of genes for both complete genome data sets



B. Homologues used for phylogenetic analyses

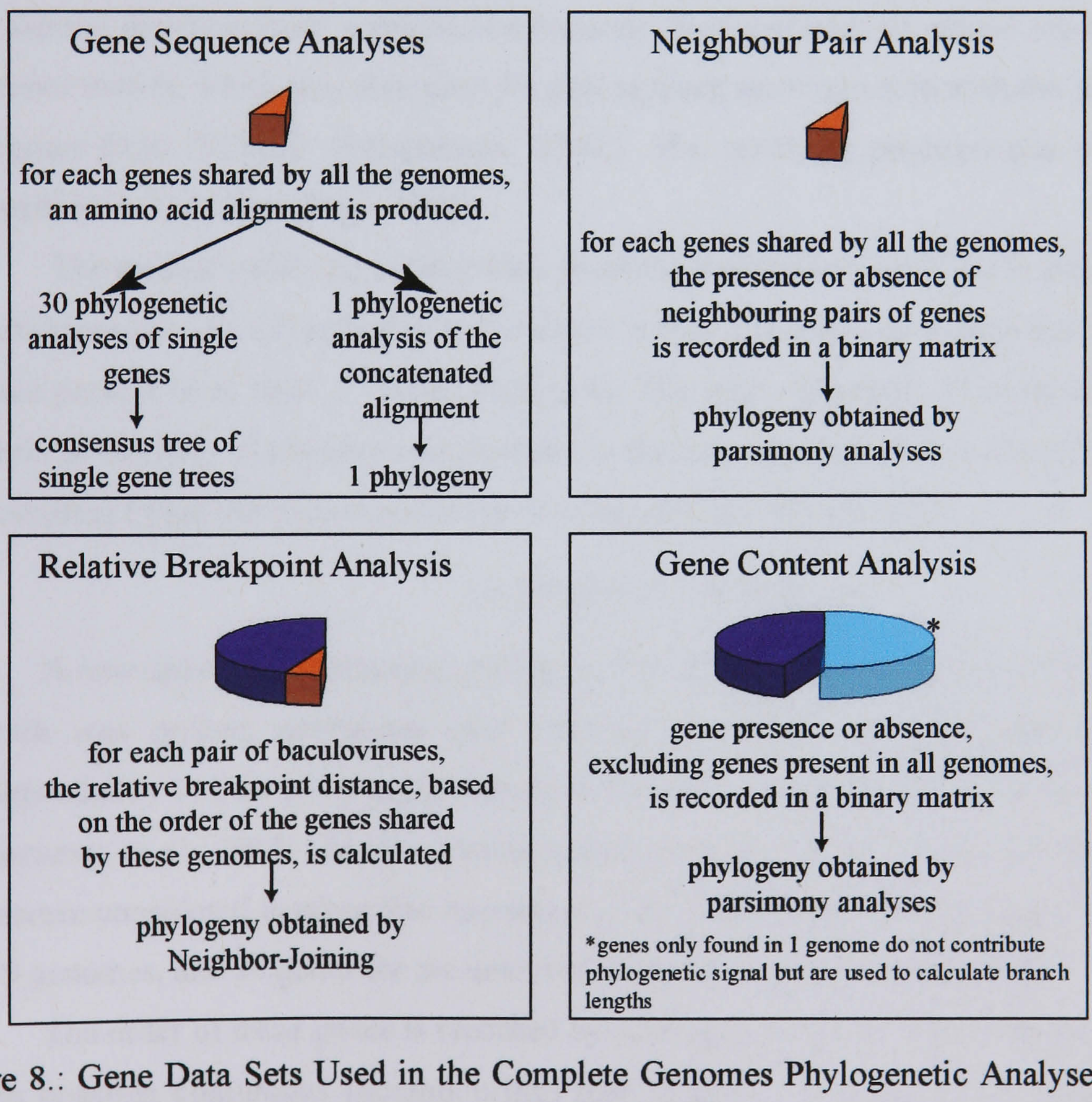


Figure 8.: Gene Data Sets Used in the Complete Genomes Phylogenetic Analyses. A. Classes of baculovirus genes B. Homologue groups used for the different phylogenetic methods.

(1) Relative Breakpoint Distances

The first phylogenetic methodology based on gene order is a modification of the breakpoint distance analysis method of Blanchette *et al.* (1999), originally described for the analysis of mitochondrial gene order. Breakpoints are the abstract points on one genome that need to be drawn to represent the breaks necessary to recreate the gene order of another genome. The breakpoints between two genomes are counted. The number obtained is subsequently divided by the number of genes in common to the pair of genomes, to yield a relative breakpoint distance. This modification was implemented to compensate for bias in the calculated distances due to differences in genome length. Without correction, comparisons between small genomes would give shorter distances simply because they have fewer genes in common. Calculation of the relative breakpoint distances from pairwise comparisons of all genomes examined resulted in a distance matrix, which was then used for phylogenetic reconstruction with the Neighbor program from PHYLIP (Felsenstein, 2000.). The resulting phylogenetic tree was visualised in TreeView (Page, 1996).

The method makes no assumptions about the mechanisms involved in genome rearrangements. An advantage of this method is that it is based on a large dataset – all genes present in at least 2 viruses (figure 8). The main drawback of using distances matrix as the base of phylogenetic analyses is that statistical methods such as bootstrap resampling to test the robustness of the resulting tree are not available.

(2) Neighbour Pair Analysis

A new approach to inferring phylogeny was also developed from gene order data, which was termed neighbour pair analysis. A similar method was proposed independently to infer phylogenies based on the order of the genes in bacterial operons (Tamames et al., 2001). In this approach only the order of the genes common to all genomes considered is taken into account (the order of 63 genes for the analyses based on 9 genomes, and 30 genes for the analyses based on 13 genomes; figure 8).

The order of these genes is recorded by scoring the presence (1) or absence (0) of each possible contiguous (neighbouring) pair of genes in each genome. The binary matrix created in this way is then used to reconstruct phylogenies using maximum parsimony in PAUP* (Swofford, 2001.). Neighbour gene pairs resulting in constant characters (present in all genomes or absent from all genomes) were not taken into account, as they result in parsimony non-informative characters. For the analyses based

on 9 genomes, it was possible to perform exhaustive searches to find the most parsimonious trees. In the case of the 13 taxa analyses this was no longer possible. Heuristic searches with 10 random additions were therefore performed instead.

Similar to breakpoint analysis, neighbour pair analysis is independent of the mechanism of gene rearrangements. It has the advantage that it allows the binary encoding, which can then be analysed by maximum parsimony. A second advantage of this method is that the binary encoding of conservation of gene order allows for the robustness of the data and phylogenetic signal to be assessed statistically. The disadvantage is that the matrix comprises a limited number of characters because it is only based on the order of genes common to all genomes (figure 8).

e) Phylogenetic Inference Based On Gene Content

When all the gene homology and orthology relationships have been ascertained, each genome can be scored for the presence (1) or absence (0) of each baculovirus gene. This generates a binary gene content matrix, which can be used to reconstruct a phylogenetic tree using maximum parsimony in PAUP* (Swofford, 2001.). Genes present in only one species provide no information to infer phylogenetic relationships between the viruses, but are used to calculate the branch lengths of the tree (figure 8). For the analyses based on 9 genomes, exhaustive searches for the most parsimonious trees were done. In the case of the 13 taxa analyses, heuristic searches with 10 random additions were performed.

f) Testing Phylogenetic Robustness

The robustness of the tree topologies obtained for each data sets were tested using a variety of approaches. If a number of different methods give the same tree, one is more inclined to believe that it represents the true evolutionary relationships of the species considered.

A distance base tree building method, Neighbor-Joining using median distances in PAUP* (Swofford, 2001.), was employed to reconstruct phylogenies with the combined gene sequences, neighbour pair and gene content data sets. This allows for a rapid comparison of the tree topologies obtained from different methods, like maximum parsimony.

It is not possible to use maximum likelihood (ML) methods in PAUP* to reconstruct phylogenies based on any of the 3 kinds of genome data sets. The reconstruction of phylogenies from amino acid sequence data using ML algorithms is not implemented in the programme. For the binary genome data sets, there are no

models of baculovirus genome evolution, so it is not possible to employ ML methods on these data sets either.

Also, because of the size of the sequence data set it would take unreasonable amounts of computing time to search for the best tree solely by ML. It is more practical to have a suite of trees and to evaluate using ML methodology, which of these is more likely to be the true tree. This was performed in the software package PAML (Yang, 1997) for the combined 63 genes data sets and 9 genomes. This programme was also used to compare the tree topologies of each data subsets using the Shimodaira-Hasegawa (SH) test (Shimodaira & Hasegawa, 1999). This test allows comparison of tree topologies and to discriminate those that are incompatible with each data set. It is the most appropriate test to use to assess the compatibility of suboptimal tree topologies in relation to individual data sets (Whelan et al., 2001).

For the binary data sets, neighbour pair analysis and gene content analysis, alternative topologies were assessed using the Kishino/Hasegawa test (KH test, (Kishino & Hasegawa, 1989)) implemented in PAUP*. This test, as it is implemented in PAUP, is not as appropriate as the SH test, because it is a 2-tail test instead of a 1-tail test (Whelan et al., 2001). But it is the only available for binary characters so far. A 1-tail test should be applied because the hypothesis tested is: "Is tree B a possible evolutionary hypothesis for the data set that had as best tree, tree A?" A 2-tail test tests the hypothesis: "which is the best tree A or B for the data set, when none of the trees are *a priori* the best tree?" Using a 2-tail instead of a 1-tail test could lead to the rejection of trees that could have been accepted. In other words, it is less conservative.

For all the data sets, bootstrap analyses were performed to evaluate the robustness of the tree topologies using all possible alternative tree building methods.

2. Results of the Phylogenetic Analyses Based On 9 Complete Genome Sequences

a) Gene Sequence Phylogenies

Comparison of the 9 genomes from lepidopteran baculoviruses included in the study revealed that they share 63 genes (*sod* (*ac31*) + 62 genes of table 1). This number has decreased to 62 for the lepidopteran baculoviruses with the addition of SpltNPV and EppoMNPV genomes that do not have a *sod* gene homologue (Hyink et al., 2002, Pang et al., 2001). It is likely to decrease further as more lepidopteran baculovirus genomes

become available. Phylogenetic trees were generated for each of these 63 shared genes resulting in 32 different tree topologies (figure 9). Most of the topological variation was in the arrangement of the GVs and in the monophyly and arrangement of the group II NPVs. The majority rule consensus tree of the most parsimonious tree for each gene (figure 10A) shows that most gene phylogenies support the NPV - GV division, and the subdivision of the NPVs into two groups.

The alignments of the 63 conserved genes were also combined and phylogenies reconstructed based on this combined alignment. This analysis yielded a single most parsimonious tree, which had high bootstrap support (figure 10B). Seven individual gene phylogenies (*ac22, ac81, ac119, ac142, ac145, lef8, lef9*) had this topology. Furthermore, SH tests showed that most individual gene phylogenies are compatible with this topology (table 3), the only exception being *odv-e66*.

The *odv-e66* gene yielded a tree that was incompatible with all individual trees and the genome trees (table 3). The phylogeny of *odv-e66* (figure 11A) only agrees with the consensus tree (figure 10A) in the arrangement of the group I NPVs. The rest of the tree suggests a complex history, possibly including several duplications, horizontal transfers and gene losses. The presence of a second copy of *odv-e66* in SeMNPV provides independent evidence for duplication (IJkel et al., 1999), suggesting that the present phylogeny could be based on paralogues instead of orthologues. The *odv-e66* gene encodes a structural protein present in the envelope of occluded virions. Understanding its complex evolutionary history might provide clues to its precise role in the virus life cycle.

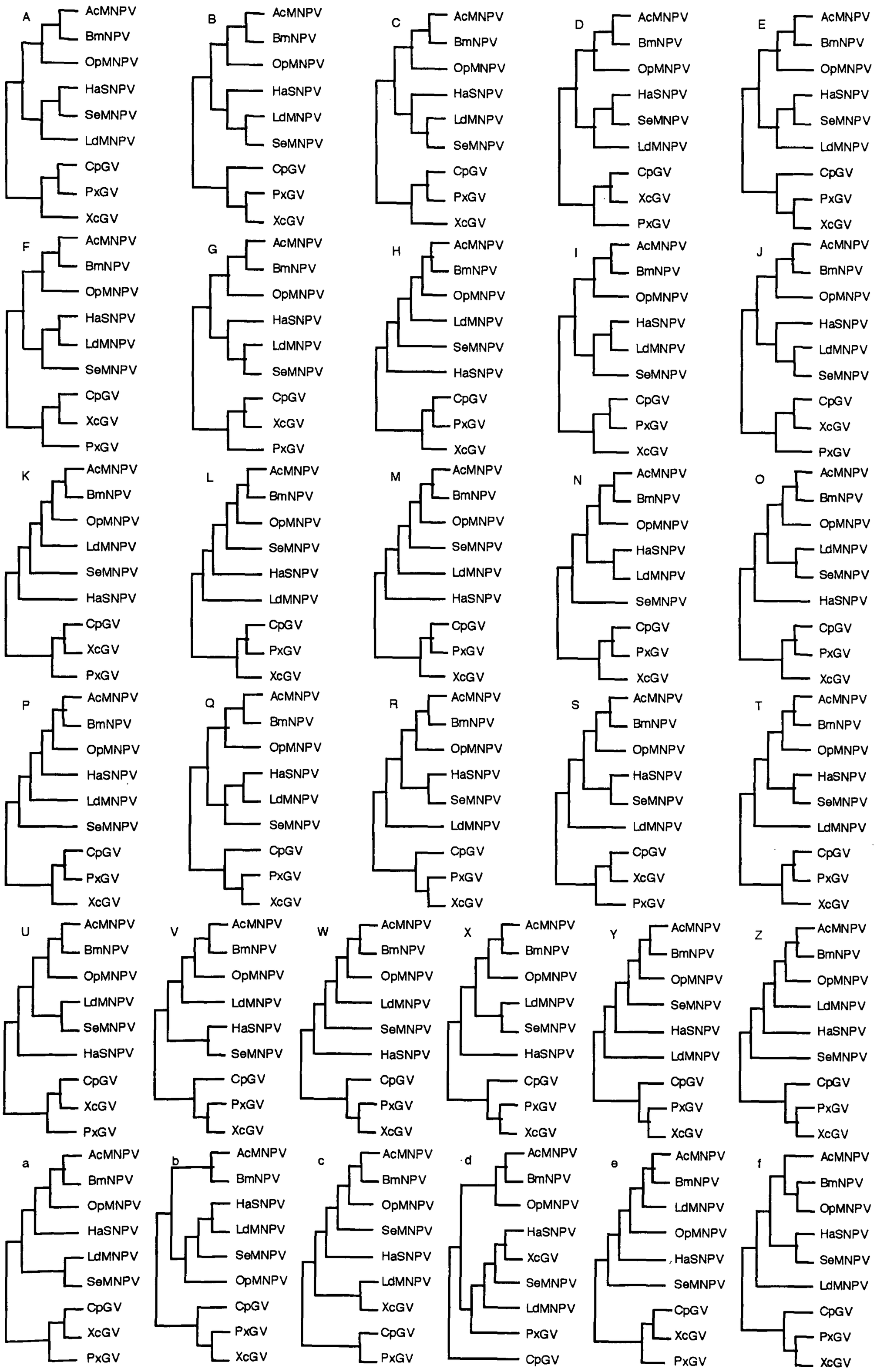


Figure 9: MPT Topologies obtained from the phylogenetic analyses of the 63 genes shared by the 9 baculovirus genomes and for the combined gene alignment, gene order, and gene content data sets. Table 3 shows which data set gave rise to each tree

Table 3: Tree Topologies Comparison - SH and KH tests

Data Set	Size †	% inf ∫	Tree Topologies																														
			A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	a	b	c	d	e	f
38.7K	422	63.7												+																		**		
38K	382	59.2		+																											**	**	**	
39K/pp31	392	59.2		+																											**	**	**	
ac22	439	51.7	+																											**	**	**	**	
ac23	770	77.9									+																			**	**	**	**	**
ac29	142	42.3		+																											**		**	
ac38	272	52.9		+																											**	**	**	
ac53	160	72.5		+																											**	**	**	**
ac66	1069	68.6													+																**	**	**	**
ac68	201	53.7																				+										**	**	
ac75	156	76.3			+																										**	**	**	
ac76	88	76.1						+																							**	**	**	
ac78	169	50.3																+															**	
ac81	267	54.3	+																											**	**	**	**	
ac82/tlp20	265	66.4				+																										**	**	**	
ac92	285	63.9							+																							**	**	**	
ac93	174	66.1			+																											**	**	**	
ac96	193	65.8							+			+																				**	**	**	
ac106	380	47.9																						+								**	**		
ac109	468	56.2														+																**	**	**	
ac110	61	75.4											+																		**				
ac115	227	65.6													+																		**	**	
ac119	583	67.6	+																													**	**	**	
ac142	520	65.2	+																													**	**	**	**
ac145	102	65.7	+																													**	**		
ac146	217	81.1															+															**	**		
alk-exo	517	59.0																										+			**	**	**		**
dbp1	365	72.6					+																									**	**		
dnapol	1200	63.4					+																									**	**	**	
fgf	456	70.8																													+	**	**		
fp/25K	249	50.6																		+													**		
gp41	421	55.1														+																**	**	**	
helicase	1349	73.4		+																												**	**	**	**
ie1	753	71.8						+																								**	**	**	
lef1	314	58.0								+																					**	**	**	**	**
lef2	317	53.3											+																			**	**		
lef3	463	75.4																					+									**	**		**
lef4	553	63.1																			+											**	**		
lef5	320	59.1				+																										**	**		
lef6	197	71.6				+																										**	**		
lef8	942	53.0	+																													**	**	**	
lef9	532	50.6	+																													**	**	**	**
lef11	193	50.3																+														**	**		
me53	523	62.1																	+													**	**	**	
odv-e18	111	56.8							+																							**	**	**	
odv-e25	242	64.5									+																					**	**	**	
odv-e56	393	64.9								+																						**	**		
odv-e66	800	59.3	**	**	**	**	**	**	**	**	**	**	**	**	**	**	**	**	**	**	**	**	**	**	**	**	**	**	**	**	**	**	+	**	**
odv-ec27	312	73.7						+																								**	**	**	
p12	134	82.1																							+							**	**		
p40	424	77.4							+			+																				**	**	**	
p45	421	67.9			+																											**	**	**	
p47	488	51.0																											+	**	**	**	**	
p6.9	117	29.9												+																			**		
p74	785	53.6				+																										**	**	**	**
p95	1096	55.5			+																											**	**	**	**
pk1	329	65.7					+																									**	**	**	
polh	248	47.6																														**	**		+
sod	160	46.3																														**	**	+	
ubi	172	20.3																														+			
vlf1	438	63.2																										+				**	**		
vp1054	632	40.8															+															**	**	**	
vp39	416	61.3																									+					**	**	**	**
all 63 genes	25788	61.7	+												**		**											**	**	**	**	**	**	**	**	**
gene content	417	32.6						*	+						*		*											*			*	*	*	*	
gene order	104	70.2				+																						*			*	*	*	*	*

† Size of the aligned sequences, ∫ percentage of informative characters, + Most parsimonious tree topology, * Tree topology significantly different to the best topology as found by the KH test performed on the binary data sets in PAUP, ** Tree topology significantly different to the best topology as found by the SH test performed on the AA sequence data sets in PAML.

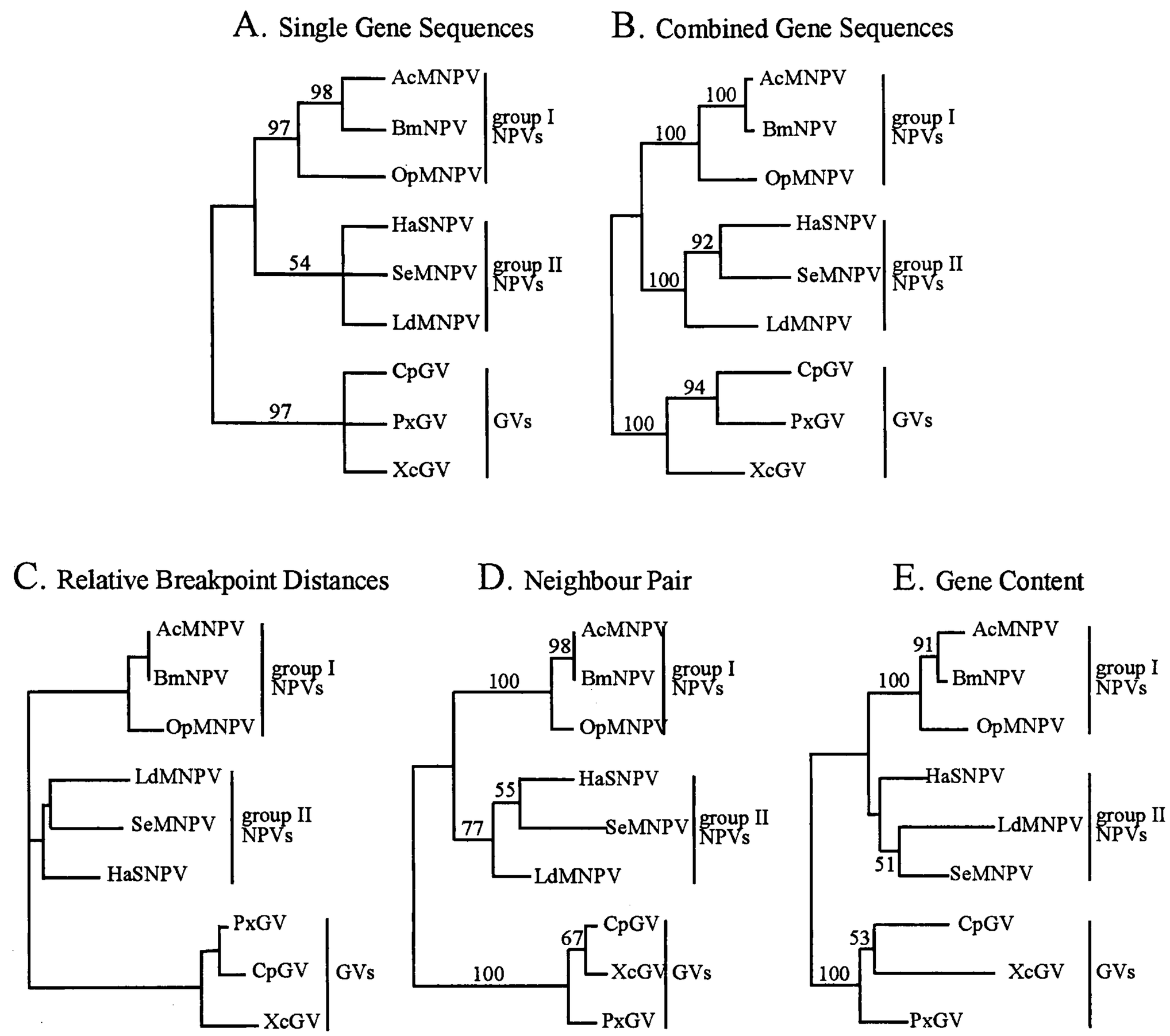


Figure 10: Phylogenies Based on the Complete Genomes of 9 Baculoviruses. A. Majority rule consensus tree of the most parsimonious trees obtained from the separate phylogenetic analysis of each of the 63 common genes; numbers above each branch indicate the percentage of individual gene trees possessing that branch. B. Most parsimonious tree based on an analysis of the combined sequences of the 63 common genes. C. Neighbor-joining tree based on relative breakpoint distances. D. Most parsiminious tree generated by neighbour pair phylogenetic analysis. E. Most parsimonious tree generated by the gene content phylogenetic analysis. Trees are rooted using the GVs as sister group to the NPVs. With the exception of A, the numbers indicate the percentage bootstrap support for each node, from 1000 replicates.

Of the 63 common genes, the only other gene whose phylogenetic tree disagrees, with strong bootstrap support, with the consensus tree (figure 10A) is the *polyhedrin* gene. The *polh* based phylogeny consistently and strongly places AcMNPV at the base of the group I NPVs (figure 11B), suggesting a horizontal transfer of the *polh* gene in the AcMNPV lineage, as previously noted (Clarke et al., 1996). The otherwise low

bootstrap scores for this tree reflect the weak phylogenetic signal in *polh* amino acid sequence alignments. Great caution should therefore be taken when interpreting phylogenies based solely on this gene.

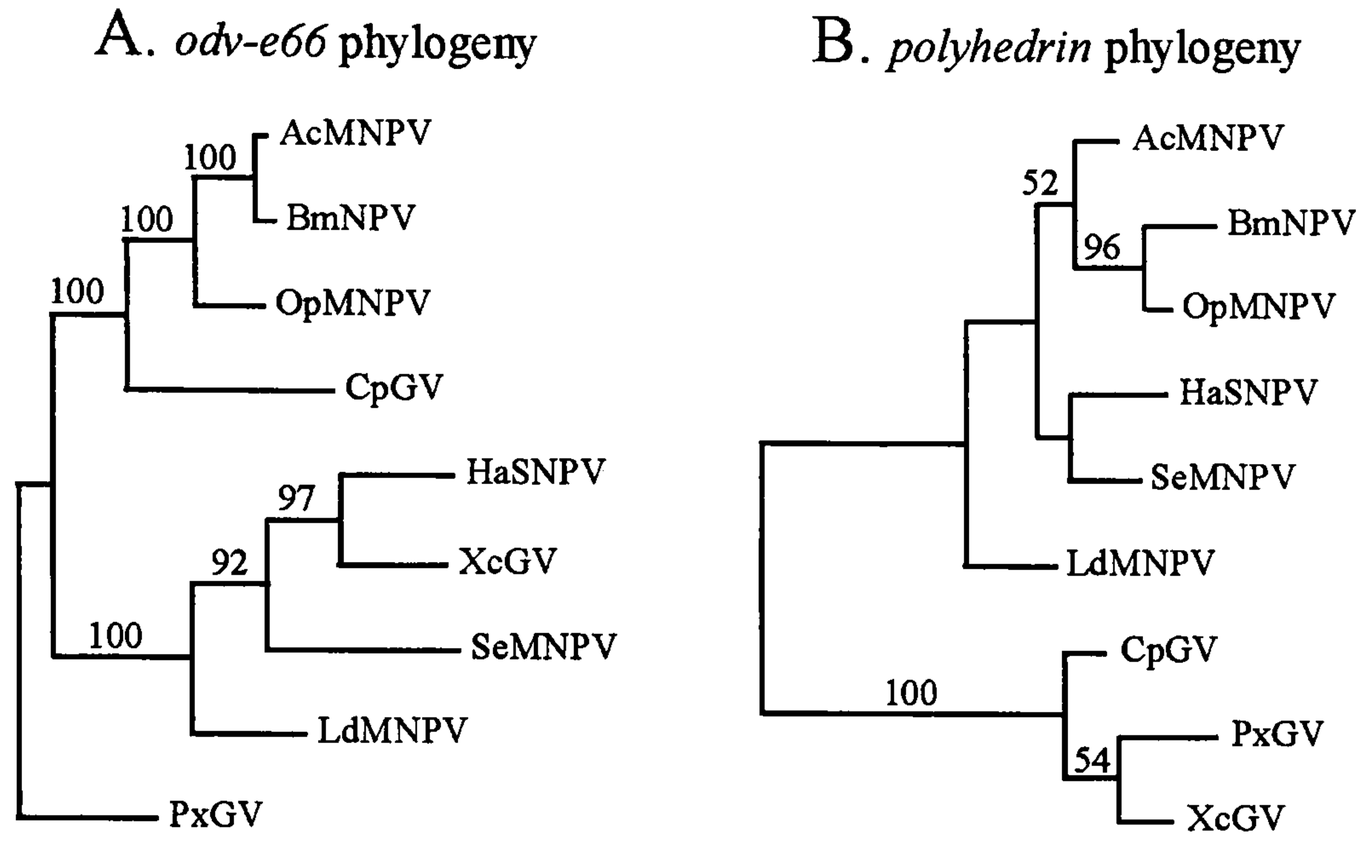


Figure 11: A. *odv-e66* and B. *polyhedrin* gene phylogenies. MPTs, percentage of bootstrap support (1000 replicates) greater than 50% are shown. The trees are rooted using the GVs as sister group to the NPVs.

b) Gene Order Phylogeny

Two approaches were used to provide a measure of the difference in synteny (i.e. gene order) between baculovirus genomes. First, a matrix of relative breakpoint distances was compiled based on pairwise comparisons of all the genomes (table 4). The distance tree generated from this matrix (figure 10C) differs from the combined gene tree (figure 10B) in the relationships among the group II NPVs, but is consistent with the majority rule consensus tree in Fig 1a. Second, a binary matrix was compiled recording the presence of conserved neighbouring gene pairs in each genome. This matrix was analysed by maximum parsimony. The most parsimonious tree (figure 10D) has a different topology again, differing from the relative breakpoint distance tree (figure 10C) in the relationships within the group II NPVs and the GVs, but only differing from the combined gene tree (figure 10B) in the relationships within the GVs. Furthermore, KH tests of this neighbour pair data set demonstrated that it is compatible with a total of 26 single gene tree topologies including both tree topologies in figure 10B and figure 10C (table 3).

Table 4: Relative Breakpoint Distance Matrix.

	AcMNPV	BmNPV	OpMNPV	EppoMNPV	LdMNPV	SeMNPV	SpltNPV	HaSNPV	HzSNPV	PxGV	XcGV	CpGV	CuniNPV
AcMNPV		0 /135	12 /127	75 /119	35 /96	37 /101	34 /94	30 /100	30 /100	37 /68	45 /77	38 /71	26 /31
BmNPV	0		11 /123	75 /118	37 /97	34 /99	34 /94	30 /100	30 /100	37 /68	42 /74	38 /71	26 /31
OpMNPV	0.09	0.09		0 /125	40 /99	40 /101	42 /96	29 /97	29 /97	38 /68	43 /74	43 /76	26 /30
EppoMNPV	0.04	0.04	0		28 /90	32 /97	36 /92	30 /97	30 /97	36 /67	40 /74	38 /72	36 /30
LdMNPV	0.36	0.38	0.40	0.31		25 /102	34 /97	23 /95	23 /95	29 /68	42 /78	36 /75	24 /30
SeMNPV	0.36	0.34	0.40	0.33	0.25		34 /106	22 /102	22 /102	34 /68	38 /73	39 /74	23 /30
SpltNPV	0.36	0.36	0.44	0.39	0.35	0.32		17 /96	17 /96	32 /68	37 /72	39 /72	24 /31
HaSNPV	0.30	0.30	0.30	0.31	0.24	0.22	0.18		0/ 132	30 /67	40 /75	34 /71	24 /30
HzSNPV	0.30	0.30	0.30	0.31	0.24	0.22	0.18	0		30 /67	40 /75	34 /71	24 /30
PxGV	0.54	0.54	0.56	0.54	0.43	0.50	0.47	0.45	0.42		12 /99	5 /99	24 /30
XcGV	0.58	0.57	0.58	0.54	0.54	0.52	0.51	0.53	0.53	0.12		18 /108	25 /30
CpGV	0.54	0.54	0.57	0.53	0.48	0.53	0.53	0.48	0.48	0.05	0.17		25 /30
CuniNPV	0.84	0.84	0.87	0.87	0.80	0.77	0.77	0.80	0.80	0.80	0.83	0.83	

Above diagonal: number of breakpoints / number of genes in common. Under diagonal: relative breakpoint distance. The distance between 2 genomes always remains the same.

c) Gene Content Phylogeny

A matrix was compiled recording the presence or absence of all baculovirus genes in each genome. Maximum parsimony analysis of this data set gave a single most parsimonious tree (figure 10E). Again, this tree separates the NPVs and GVs and resolves the NPVs into two subgroups. It differs from previous trees in the relationships among the group II NPVs and the GVs. The tree is consistent with the majority rule consensus tree (figure 10A). KH tests demonstrated that it is also compatible with 24 of the single gene trees, including all tree topologies in figure 10 (table 3).

d) How Do The Genome Based Phylogenetic Methods Fare?

Comparative genomics will become an increasingly powerful tool for inferring biological function as more genome sequences become available. However, to exploit this approach fully, it will be critical to develop methods that place the data in an appropriate evolutionary context. Baculoviruses provide here a case in point. It is essential to establish relationships among baculoviruses reliably to effectively interpret the wealth of information about the biology and evolutionary history of these viruses contained within the complete genome sequence data. The rapidly increasing availability of complete genome sequences has prompted an interest in using information other than nucleotide or amino acid sequence data for the generation of molecular phylogenies. Gene content has already been used in phylogenetic analyses of the herpesviruses, prokaryotes and eukaryotes (Fitz-Gibbon & House, 1999, Montague & Hutchison, 2000, Snel et al., 1999, Tekaia et al., 1999). Gene order has also been used to reconstruct phylogenies of the herpesviruses, animal mitochondrial genomes and bacteria. However, its use can be hindered by a lack of synteny conservation or a lack of synteny variation (Blanchette et al., 1999, Hannenhalli et al., 1995, Le et al., 2000, Tillier & Collins, 2000). Here methods based on gene order, gene content, and conserved gene sequences were evaluated for reconstructing the relationships between nine lepidopteran baculoviruses. This represented the most comprehensive analysis to date of baculovirus phylogeny.

All the approaches used agreed on the separation of the NPVs and GVs, and the division of the NPVs into groups I and II, as postulated in Zanotto *et al.* (1993), and Bulach *et al.* (1999). They all also resolved the relationships among the group I NPVs. Relationships between viruses in the other groups were only clearly resolved by the combined gene sequence analysis (figure 10B). Several lines of evidence support this tree as the most plausible representation of the relationships between these 9 viruses. First, it is strongly supported by bootstrap analysis (> 90% support for all nodes). Second, it is based on a large dataset. It has been observed previously that with a consistent method, combining genes reduces sampling error and causes the phylogenies to converge towards the correct solution with good support (Mitchell et al., 2000). Third, although the gene order and gene content-based analyses yielded different optimal topologies, the combined gene topology was always present among sub-optimal trees that were compatible with the data (table 3). Finally, this tree was also found to be

UNIV.

the best tree when SH tests where performed for the whole data set under maximum likelihood criteria in PAML (Yang, 2000).

The fact that methods based on gene order and gene content could resolve the virus relationships into the three major groups demonstrates that such approaches do permit phylogenetic reconstruction. However, the data available from the baculoviruses that have been sequenced to date are relatively weak and prone to homoplasy. Homoplasy is defined as similarity due to independent evolutionary change. This can either be due to convergent evolution (e.g. two genomes appear similar because they have independently acquired the same gene) or reversal to an ancestral state (e.g. two genomes appear similar because a gene was acquired and subsequently lost during the evolution of one, but was never present in the lineage of the other). The weakness of the data was reflected in the large number of suboptimal trees compatible with the datasets, as shown by KH tests. Gene order and gene content have the advantage of providing independent data sets from the gene sequences, with independent dynamics and rates of evolution. This is particularly true for gene content analysis, as the parsimoniously informative genes used (present in more than one genome but not present in all) do not overlap with the genes used for sequence-based analyses (present in all genomes). It is anticipated that the future addition of more species from the group II NPVs and the GVs will improve species sampling and reduce homoplastic noise, and thus provide better phylogenetic resolution using whole-genome based approaches. The approaches developed here should also prove valuable for the phylogenetic analysis of other organisms, including other large DNA viruses.

It is worth noting that, although the gene sequence data contains the strongest phylogenetic signal, only a few individual genes actually gave the best tree (*ac22, ac81, ac119, ac142, ac145, lef8, lef9*). This underlines the danger of using phylogenies based on one or a small number of genes to infer the evolution of whole genomes or species. Thus, the reconstruction of baculovirus phylogenies should <u>*ideally*</u> be based on a combined analysis of all genes conserved among all baculoviruses. Whole genome approaches based on gene content and gene order should be used to complement this analysis, as it is clear that both are phylogenetically informative and can provide additional support for the combined gene tree. As more genomes are sequenced, they should become increasingly powerful tools.

Most of the topological variation between the data sets resided within the GVs and group II NPVs. A number of factors contribute to this. First, both groups contain

one genome (XcGV and LdMNPV) that is markedly larger than the others, creating an imbalance in the character distribution for gene content and gene order resulting in a long-branch attraction effect (Felsenstein, 1978). This is most noticeable for the gene content phylogeny (figure 10E) where smaller genomes are attracted to each other at the base of their respective groups. Second, species within the groups are either too similar or too different to provide appropriate characters to resolve their relationships. Gene order data are not very informative for the GVs because of the almost identical order of the 63 conserved genes among these viruses (figure 10D). Conversely, relationships within the group II NPVs are obscured by their extensive differences in genome arrangements.

An additional pattern emerging from these data is that the monophyly of the group II NPVs is far less well supported than that of the group I NPVs or the GVs. This could indicate a sampling artefact whereby the species representing the other two groups are much more closely related than the group II species. Alternatively, it might indicate that the group II NPVs are an older group than the other two. Similarly the understanding of baculovirus evolution might change when more complete genomes become available for phylogenetic analysis.

3. Results of the Phylogenetic Analyses Based On 13 Complete Genome Sequences

a) Gene Sequences

Comparison of the 13 baculovirus genomes included in this study (table 2) revealed that they share 30 genes (table 1). The results of each of the 30 single gene analyses were extremely heterogeneous. No two single gene analyses gave the same tree. However, the majority rule consensus tree (figure 12A) demonstrates that there is a single tree underlying the different topologies. The figures above each branch show the percentage of gene trees supporting that branch. The majority of genes support the separation of CuniNPV from the lepidopteran viruses, the division between GVs and NPVs, and the separation and relationships among the group I NPVs. However, the low percentage values of branches leading to the group II NPVs, of branches within the group II NPVs and of branches within the GVs, shows that there was little agreement between genes on these relationships. Similar heterogeneity has been observed in

previous phylogenetic analyses of some of these 30 genes (Bulach et al., 1999, Chen et al., 1999, Hayakawa et al., 2000, Liu & Maruniak, 1999, Moser et al., 2001).

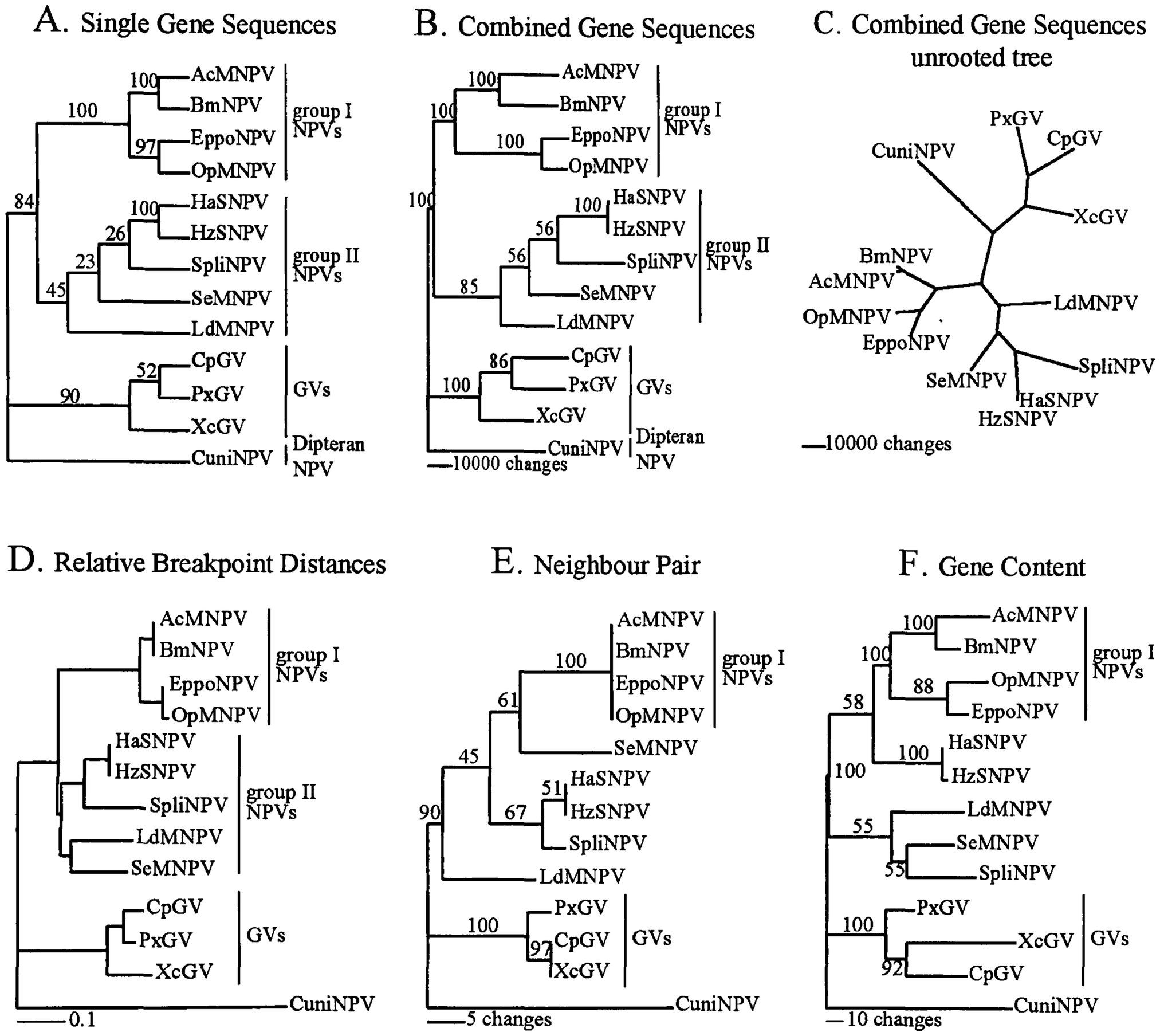


Figure 12: Phylogenies Based on the Complete Genomes of 13 Baculoviruses. A. Majority rule consensus tree of the most parsimonious trees obtained from the separate phylogenetic analysis of each of the 30 common genes; numbers above each branch indicate the percentage of individual gene trees possessing that branch. B. Most parsimonious tree based on an analysis of the combined sequences of the 30 common genes. C. Unrooted version of B. D.Neighbor-joining tree based on relative breakpoint distances. E. Most parsiminious tree generated by neighbour pair phylogenetic analysis. F. Most parsimonious tree generated by the gene content phylogenetic analysis. Trees are rooted with CuniNPV. With the exception of A, the numbers indicate the percentage bootstrap support for each node, from 1000 replicates.

As expected in the case of weak single gene data sets, the concatenation of all 30-gene alignments gives a much stronger phylogenetic signal. The combined gene tree

(figure 12B-C) has the same topology as the majority rule consensus tree (figure 12A), but has high bootstrap values, and shows that 4 main groups can be distinguished within the baculoviruses. The group I NPVs, which includes AcMNPV, BmNPV, OpMNPV and EppoMNPV, is the best-supported and most highly resolved group. There is strong support for the existence of the group II NPVs, comprising HaSNPV, HzSNPV, SeMNPV, SpltNPV, & LdMNPV. However, relationships within the group are not well supported. Both groups of lepidopteran NPVs are more closely related to each other than they are to either the GVs or to CuniNPV. The lepidopteran NPVs and GVs also group strongly together to the exclusion of the dipteran virus. This corroborates previous results where CuniNPV was identified as a divergent baculovirus based on morphological data and a DNA polymerase phylogeny (Moser et al., 2001).

b) Gene Order

The neighbour pair analysis tree (figure 12D) shows clear divisions between CuniNPV and the lepidopteran baculoviruses and also between the NPVs and the GVs. However, it is poorly resolved and fails to establish relationships within the group I NPVs. This is mostly due to the inclusion of the dipteran virus, which reduced the common gene pool from 62 between all lepidopteran baculoviruses (Hyink et al., 2002) to 30 genes. In fact, there are no differences in the order of these 30 common genes among the group I NPVs, between HaSNPV and HzSNPV, or between CpGV and XcGV, as shown by the null branch lengths on the tree. Furthermore the neighbour pair analysis fails to recover the group II NPVs and the low bootstrap values indicate that branching in this area of the tree is weakly supported. In contrast, the relative breakpoint distance tree (table 4, figure 12E) is well resolved. It recovers the 4 main groups of baculoviruses and its topology only differs from the 30-gene tree (figure 12E) in the grouping of SeMNPV and LdMNPV within the group II NPVs.

c) Gene Content

The gene content tree (figure 12F) again shows a clear division between CuniNPV, the GVs and the lepidopteran NPVs. High bootstrap values show that relationships within the group I NPVs and the GVs are strongly supported. However, compared to the 30-gene trees and the relative breakpoint tree (figure 12A-C, E), this analysis does not support the group II NPVs. This is due to the fact that only a few genes are specific to the group II NPVs. As observed for bacterial genome analysis (Snel et al., 1999), a limitation of gene content analysis is that, regardless of their relatedness, larger genomes are more likely to share more genes and thus to appear

more closely related i.e. they are prone to long branch attraction artefacts. An example is the apparent relationships within the GVs. The gene content analysis places PxGV, which has a much smaller genome than the other 2 GVs, basal to the group. However the 30-gene tree and other observations show that in fact it is more closely related to CpGV, and that XcGV should be at the base of the group. Nonetheless, the gene content analysis managed to recover the main relationships between the baculoviruses.

d) Do The Addition Of More Taxa Affects The Genome Based Phylogenetic Methods?

The addition of more complete genomes has been a mixed blessing for the complete genomes analyses. The addition of 4 genomes including that of the mosquito virus CuniNPV has in many ways changed the shape of various genome data sets, and therefore some of the resulting phylogenies.

This is most striking in the case of the neighbour pair analysis, where the arrangements of the 30 remaining genes in common are not sufficient to resolve the relationships within the GVs or the NPVs. From the previous neighbour pair data set, the number of informative characters dropped from 73 to 42. This would explain why the neighbour pair data set failed to resolve the relationship between the 13 baculoviruses.

In comparison, the relative breakpoint distance analysis fared pretty well, all major groups being recovered. The distance measure of gene order is not modified when more taxa are added, making this method quite robust.

The gene content data set was already prone to homoplasy in the 9 genomes analysis. The addition of more taxa did not seem to reduce the homoplasy. On the contrary, the noise seems to have increased more than the phylogenetic signal as this time the data set failed to recover the group II NPVs.

Although the combined data set of the genes in common to all genomes has been reduced in size by more than half the genes, the overall impression is that the signal with more taxa and less sequence is equally good as in the 9 genomes analysis.

In summary, the methods that appeared to be the most robust in the 9 genomes analyses remained the most robust. The combined alignment of all the genes in common to the genomes to analyse should still be the preferred method to reconstruct a phylogeny based on complete baculovirus genomes. The relative breakpoint analysis also seems quite robust and provides another perspective of genome evolution because it is based on genome organisation.

4. A Phylogeny Of The Baculoviridae

The different methods to reconstruct baculovirus phylogenies based on complete genomes each have positive points in their own right. For the gene order analyses, as well as resolving the basic relationships of the baculoviruses, the branch lengths of the trees provide an approach to visualizing divergent genome arrangements. As noted, they are limited in their resolving power when genome orders are too similar or too divergent. However, they could prove extremely useful in intermediate cases. The gene content analyses are conceptually simple but also have limited power. However, they are extremely useful for addressing questions of genome evolution. The combined analysis of all the genes that are common between all the genomes is the most robust approach for infering baculovirus phylogenies. The proposed phylogeny of the 13 baculoviruses has 4 distinct groups, the dipteran virus (CuniNPV), the GVs, and the 2 groups of lepidopteran NPVs (figure 12B). It is worth noting that the underlying branching pattern of the phylogeny obtained from the 13 sequenced baculoviruses is the same as the phylogenies obtained from 9 genomes (figure 12B) or 10 genomes (Hyink et al., 2002). Thus, it is reasonable to assume that it represents a robust and plausible phylogeny of these viruses. In addition to its unusual characteristics (Moser et al., 2001), the genome phylogeny clearly shows that the mosquito virus CuniNPV belongs to a different group to the lepidopteran NPVs. Thus, the taxonomy of the *Baculoviridae* should be re-examined. It is no longer appropriate to include all NPVs morphotypes within a single genus *Nucleopolyhedrovirus*.

There is abundant evidence in the data analysed here of the fluid nature of baculovirus genomes. Of more than 555 genes identified in these thirteen genomes, only 30 are present in all genomes, and 282 are present in only one genome (figure 8), although it is conceivable that some of these might represent highly diverged homologues not recognised by current comparison methods. Similarly, analysis of the gene order data points to frequent gene rearrangements in the course of baculovirus evolution. For example in the neighbour pair analysis, the patristic distance between AcMNPV and CpGV is 36, implying a minimum of 36 rearrangements between the 30 conserved genes since their last common ancestor.

As noted above, comparison of individual gene phylogenies to the whole genome phylogeny provides further support for horizontal transfer of genes between genomes. Despite this fluidity, it was possible to recover a single, well-supported tree that

describes the evolution of these thirteen viruses. The challenge now will be to relate biological differences to the evolutionary groups that have been highlighted, so that the features of baculovirus biology and ecology, which have driven the diversification of this virus family, can be understood.

C. Comparative genomics

Following the many genome projects, which have been completed recently, extensive efforts have been made to develop methods to analyse those sequences in a comparative and evolutionary context, to improve our understanding of genome evolution (Eisen, 2002, Koonin et al., 2000). For instance, the study of gene content has shown the extent of variation between genomes (Alm et al., 2000). Protein phylogenetic profiles have been used to infer the functions and interactions of proteins (Bork et al., 1998, Dandekar et al., 1999, Dandekar et al., 1998, Eisenberg et al., 2000). Comparative genomics has the potential to provide valuable insights into baculovirus evolution and biology. The comparison of closely related viruses can indicate which genes are involved in fine-tuning their biology. Conversely, the study of more distantly related genomes can highlight the molecular innovations characterising each group. In this section, using the phylogeny determined above based on 13 complete baculovirus genomes, some comparative analyses of these genomes are reviewed and placed in an evolutionary framework.

1. The evolution of genome content

A highly informative way to visualize the gene content data is to map it onto the optimal phylogenetic tree, revealing where gene content changes are likely to have occurred during the evolution of baculoviruses.

a) Method of gene content mapping

The evolution of genome content can be visualized by mapping the gene content data set onto a phylogenetic tree. Character changes (i.e. gene acquisition or loss) are mapped onto the trees using MacClade 4. The programme uses maximum parsimony algorithms to reconstruct character evolution on a given tree. It is therefore possible to represent the gene content changes that have happened during the evolution of the 13 genomes based on the observed current gene content and on the phylogeny (figure 12B) that is believed to illustrate their true relationships.

Each character can evolve in a number of ways on the tree. The characters only have 2 possible states - either present in the genome (1) or absent (0). Therefore the changes in character state can be either from absent to present (0>1) or from present to absent (1>0) at any one time (figure 13). However only the most parsimonious

explanation is mapped on the tree. MacClade will only draw the minimum number of character changes that are necessary to explain the current pattern of presence or absence for each gene in the genomes that we observe.



Figure 13: Patterns of gene content evolution, possible mapping of gene acquisition and losses on a phylogeny.

(i.) (0) Constant, absent: these characters are not in the database, they correspond to genes that have no homologues in the 13 baculoviruses genomes

(ii.) (1) Constant, present: these correspond to the 30 genes that are present in all the genomes. They are represented at the root of the tree and assumed to be the genes present in the baculovirus ancestor.

(iii.) (1>0) One loss: this represents the loss of a gene that was present at the root of the tree. As we do not know the gene content of the baculovirus ancestor, there is always the possible alternative that the gene was not present at the root of the tree but acquired later. To reduce the ambiguity, it was decided to ignore this possibility for representation purposes. The ancestor harbours only the genes common to all genomes and none that would be lost later. This allows a clearer depiction of the genes that define each group of viruses, even though it implies that the ancestral genome had very few genes and that a large number of genes were acquired at the beginning of baculovirus diversification. This assumption is equivalent to using an outgroup, defined as a non-baculovirus, that would not contain any baculovirus gene homologues. This orientates the evolution of characters from 0 to 1.

(iv.) (0>1) One acquisition: This represents the acquisition of a gene at a certain point in baculovirus evolution. It can be on a branch of the tree that leads to a

single genome and then represents the acquisition of a unique gene by one virus. It can be on a branch that leads to a group of species and then represents the acquisition of a gene specific of one clade and can define the group. It is always the alternative hypothesis to a gene that could have been present in the ancestor and lost at one point during the evolution.

(v.) (0>1>0; …) One acquisition of a gene that is then lost one or several times

(vi.) (0>1; 0>1; …) Several acquisitions of the same gene

b) Evolution of baculovirus gene content based on 13 genomes

The acquisition and losses of 555 baculovirus genes were retraced on the 30-gene phylogeny. The tree in figure 14 summarises the most parsimonious hypothesis of changes in gene composition that have occurred during the evolution of these 13 baculovirus genomes. The details of which genes have been acquired when are listed in table 5.

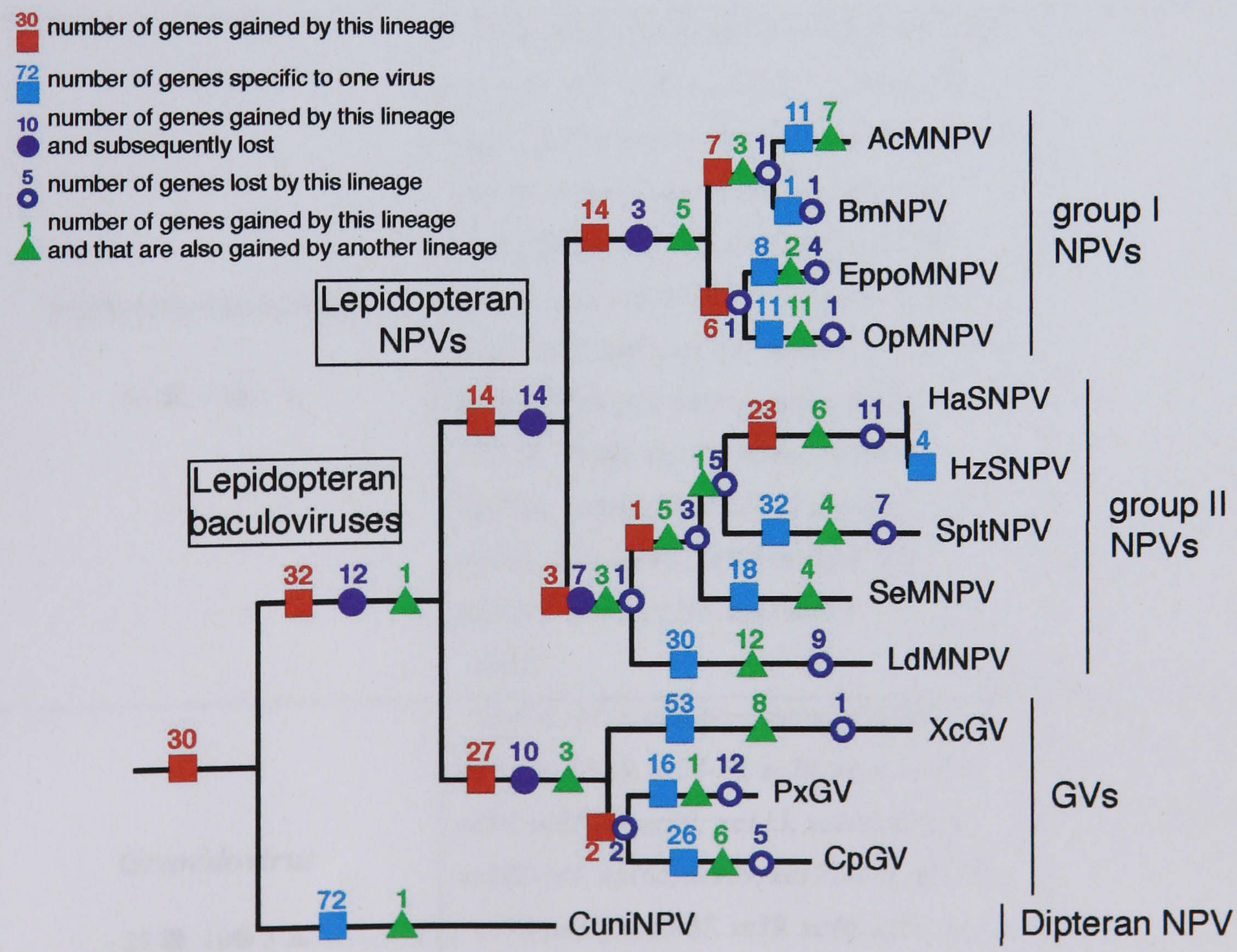


Figure 14: Gene content mapping. Most parsimonious hypothesis of changes in gene content during baculovirus evolution. The gene content data set is mapped onto the tree derived from the analysis of combined sequences of the 30 shared genes. Genes have been subdivided into 5 groups based on their patterns of acquisition and loss.

Table 5: Details of the gene acquisitions and losses which have occurred during baculovirus evolution

Description of the lineage Summary of gene acquisitions&losses *	Acquisitions	Losses
Baculoviridae 30 ■	*lef-1 (ac14), lef-2 (ac6), dnapol (ac65), helicase (ac95), p47 (ac40), lef-8 (ac50), lef-9 (ac62), vlf-1 (ac77), lef-4 (ac90), lef-5 (ac99), ac23 (ld130), gp41 (ac80). odv-ec27 (ac144), odv-e56 (ac148), p6.9 (ac100), p74 (ac138), vp91/p95 (ac83), vp39 (ac89), vp1054 (ac54), alk-exo (ac133), 38K (ac98), ac22, ac68, ac81, ac92, ac96, ac109, ac115, ac119, ac142*	
CuniNPV 72 ■ 1 ▲	*cn2, cn3, cn6, cn7, cn9-13, cn15-17, cn19-22, cn27, cn28, cn31, cn34, cn36, cn37, cn39-44, cn47-53, cn55-57, cn60-68, cn70-72, cn76-86, cn92-94, cn97-101, cn103, cn105, cn107 p35(cn75/ac135)*	
lepidopteran bacoloviruses 32 ■ 12● 1 ▲	*dbp1 (ac25), lef-3 (ac67), ie-1 (ac147), me53 (ac139), Pp31/39K (ac36), lef-6 (ac28), lef-11 (ac37, fp25K (ac61), odv-e18 (ac143), odv-e25 (ac94), odv-e66 (ac46), pk1 (ac10), polh (ac8), fgf (ac32), ubiquitin (ac35), 38.7K (ac13), ac29, ac38, ac53, ac66, ac75, ac76, ac78, ac82, ac93, ac106, ac110, ac145, ac146, p40 (ac101), p12 (ac102), p45 (ac103) 11.3K (ac79), ac150, ac60, chitinase (ac126), cathepsin (ac127), egt (ac15), gp37 (ac64), ie0 (ac141), lef10 (ac53a), p10 (ac137), p24 (ac129), sod (ac31) ac111*	
Granulovirus 27 ■ 10● 3 ▲	*iap5 (xc137), metalloproteinase (xc40), xc2, xc7-8, xc17-19, xc25-26, xc29, xc34, xc47, xc54, xc85-86, xc90, xc113, xc116, xc136, xc142-143, xc165, xc169, xc172-173, xc178 xc30, xc33, xc36-37, xc39, xc46, xc66, xc124, xc170, xc176 DNAligase(xc141), helicase2 (xc146), p13(xc43)*	

CpGV + PxGV 2 ■ 2 o	*cp62, cp130*	*ac60, ac111*
CpGV 26 ■ 6 ▲ 5 o	*iap-cp1 (cp94), ptp2a (cp98), cp6, cp12, cp21, cp25-28, cp32, cp36, cp40, cp43-44, cp51, cp61, cp97, p70, cp72, cp78, cp109-110, cp124, cp129, cp139, cp142* *iap3, ld138, pe38, ptp2, rr1, rr2a*	*ie0, p10, xc39, xc124, xc176*
PxGV 16 ■ 1 ▲ 12 o	*px1, px3, px18-19, px22, px27, px33, px39, px48, px58, px68, px77, px81, px105, px111, px119* *ac91*	*11.3K (ac79), chitinase, cathepsin, gp37, lef10, xc30, xc33, xc36-37, xc46, xc66, xc170*
XcGV 53 ■ 8 ▲ 1 o	*enhancin1 (xc150), enhancin4(xc166), xc104-106, xc108, xc114-115, xc117, xc128-129, xc138, xc151, xc153, xc155-158, xc161-164,xc167-168, xc177, xc179, xc181, xc20, xc22-24, xc28, xc38, xc4, xc41-42, xc44, xc48-49, xc57, xc59, xc6, xc61-65, xc69-74* *ac113, conotoxin2, enhancin2-3 (xc152-154), he65, p94, xc31, xc83*	*egt*
Lepidopteran NPVs 14 ■ 14●	*arif1 (ac21), iap2(ac71), orf1629(ac9), pkip (ac24), pp34(ac131), vp80 (ac104), ac108, ac17,ac19, ac26, ac34, ac55, ac57, ac59* *ac4, ac11, ac18, ac43, ac51-52, ac56, ac69, ac117, ac120, cg30 (ac88),gp16 (ac130), iap3, p26 (ac136)*	
Group II NPVs 3 ■ 7● 3 ▲ 1 o	*ld55, ld129, parg (ld141)* *rr2b, orf4PE, ld111, ld124, ld127, ld142, ld144* *rr1, dUTPase, ld138,*	*11.3K*
LdMNPV 30 ■ 12 ▲ 9 o	*Ld4-13, ld24-26, ld28, ld31, ld34, ld49, ld52, ld59, ld69, ld77, ld109, ld121, ld126, ld132-134, ld152, ld162-163* *ld 26, ld135, enhancin2-3, hrf1, conotoxin1-2, rr2a, ac12, ac63, DNAligase, helicase2*	*p24, gp16, ac117, ac120, ac43, ac51, ac56, ac69, cg30*
SeMNPV, SpliNPV, HzSNPV, HaSNPV	*haor* *p13, se22-24, se68*	*ac4, ac11, ac111*

1 ■ 5 ▲ 3 o		
SeMNPV 18 ■ 4 ▲	*odve66(2), p26(2), se5, se17-18, se20-21, se31, se39-40, se44, se83, se85-86, se116-117, se121-122* *ac33, ac44, p94, ptp2*	
SpliNPV, HzSNPV, HaSNPV 1 ▲ 5 o	*lef12*	*Ld111, ld127, ld142, ld144, orf4PE*
SpltNPV 32 ■ 4 ▲ 7 o	*sl5-7, sl9, sl20, sl24-26, sl41, sl43, sl47, sl56, sl58, sl60-62, sl70, sl103-104, sl106, sl108, sl111, sl115, sl117, sl126-127, sl131, sl137-141* *helicase-2, ld135, op98, p35*	*sod, p26, ac117, ac43, ac56, gp16, se68*
HaSNPV + HzSNPV 23 ■ 6 ▲ 11 o	*ha5-7, ha18, ha29-30, ha34, ha40, ha45, ha54, ha99, ha102, ha104, ha107-109, ha112, ha122, ha125, ha127, ha134-135, hz119* *ac63, ac74, he65, ld26, xc83, ac111*	*ac150, ac18, ac120, dUTPase, ld124, ld138, rr1, rr2b, se22-24*
HzSNPV 4 ■	*hz26, hz42, hz62, hz79*	
HaSNPV		
Group I NPVs 14 ■ 3 ● 5 ▲	*ac5, ac30, ac72-73, ac114, ac124, ac132, gp64 (ac128), gta (ac42), iap1 (ac27), ie2 (ac151), lef7 (ac125), odve26 (ac16), ptp1 (ac1)* *etm (ac48), ets (ac47), vp80a (ac87)* *ac44, ac74, ac91, lef12 (ac41), pe38 (ac153)*	
OpMNPV + EppoMNPV 6 ■ 1 o	*Iapop4, op4-5, op36-37, op113*	*ac52*
OpMNPV 11 ■ 11 ▲ 1 o	*Op28, op33, op66-68, opep25, opep32, p8.9* *hrf1, conotoxin1-2, rr1, rr2a, dUTPase, ld138, ac85, pcna, op98, ptp2*	*ac69*
EppoMNPV 8 ■ 2 ▲ 4 o	*ep9, ep28, ep43, ep98, ep105, ep113, ep132, ep134* *xc30, xc31*	*sod, ac150, ets, vp80a*

AcMNPV + BmNPV 7 ■ 3 ▲ 1 o	*Ac45, ac116, ac121, ac149, ac154, p43 (ac39), pk2 (ac123) p35(ac135), he65, ac63*	*iap3*
BmNPV 1 ▪ 1 o	*bm111*	*etm*
AcMNPV 11 ▪ 7 ▲	*ac7, ac58, ac70, ac84, ac97, ac107, ac112, ac118, ac140, ac152, pnk (ac86) ac12, ac33, ac85, ac113, conotoxin1, p94 (ac134), pcna (ac49)*	

* 7 ■ number of genes gained by this lineage, 1 ▪ number of genes specific to one virus, 3 ● number of genes gained by this lineage and subsequently lost, 1 o number of genes lost by this lineage, 3 ▲ number of genes gained by this lineage and that are also gained by another lineage.

The ancestor of all baculoviruses had at least 30 genes in its genome (figure 14, table 5). These genes show a large array of the basic functions necessary for baculovirus replication and can easily be envisaged as the essential genes that the ancestral virus could have harboured (see A, above). If a more divergent virus than CuniNPV was to be sequenced, the number present in the ancestor of baculoviruses could increase if that genome shared more genes with the lepidopteran baculoviruses or it could decrease if it shared fewer genes.

The lineage leading to CuniNPV acquired 72 genes, which have no homology with anything in the databases. It also acquired a gene that has some homology to the baculovirus antiapoptotic gene *p35* (Afonso et al., 2001). This homology is extremely low at the sequence level and further investigation on the function of this gene will be required to validate fully the nature of this sequence homology.

The ancestor of the lepidopteran viruses (figure 14) acquired 45 genes, 32 of which are still present in all lepidopteran virus genomes sequenced so far (tables 1, 5), 12 of which have been subsequently lost in at least one lineage, and one of which (*ac111*) seems to have been lost in some lineages and subsequently reacquired (table 5). One can speculate that the acquisition of these 45 genes might have been crucial to the colonisation of lepidopteran hosts. However, it is important to remember that, since only one non-lepidopteran baculovirus has been sequenced to date, the evidence that these genes are specific to lepidopteran viruses is weak. Interestingly, several of the 12 genes that are lost in some lineages had already been classified as auxiliary genes: *egt, chitinase, cathepsin, sod* and *p10* (O'Reilly, 1997). This supports the idea that these genes are beneficial for viral replication in Lepidoptera but not essential.

The gene content mapping tree indicates that 27 genes specific to GVs and 14 to lepidopteran NPVs (figure 14) could underlie the biological differences of these 2 baculovirus groups. Unfortunately, potential functions have only been ascribed to 6 of the NPV specific genes and 2 of the GV specific genes. Some of the structural differences between NPVs and GVs could be associated with 3 NPV specific genes (*vp80*, *pp34/calyx*, and *orf1629*) that code for structural proteins (Funk et al., 1997, Hayakawa et al., 2000). The presence in the NPVs of *arif 1*, which is implicated in rearrangement of the cytoskeleton (Roncarati & Knebel-Mörsdorf, 1997), may be linked to differences in subcellular architecture during NPV and GV infections (Winstanley & O'Reilly, 1999). The *iap* genes are implicated in the inhibition of apoptosis (Clem & Miller, 1994) in baculovirus infected cells. It is intriguing that individual members of this gene family appear to be unique to both GVs (*iap5*) and NPVs (*iap2*). The only other GV unique gene with a potential function encodes a metalloproteinase, which is thought to contribute to the proteolysis of infected tissue (Ko et al., 2000). Further correlation between differences in gene content and biological differences of NPVs and GVs awaits the characterisation of genes of unknown function.

Seventeen genes are only found in the group I NPVs, 14 in all of them and 3 (*ets, etm* and *vp80a*) in some of them (figure 14). The best studied of these is probably *gp64*, which has been found to have the analogous function of the baculovirus F protein (*bfp*) homologues from group II NPVs, as discussed earlier (IJkel et al., 2000, Pearson et al., 2000). The *gp64* gene is related to a Thogoto virus (a tick-borne orthomyxo-like virus) glycoprotein (Morse et al., 1992). It represents the best example of non-orthologous gene displacement in baculoviruses. It has been suggested that acquisition of this gene promoted the diversification of group I NPVs because of its unique structural role, which is to facilitate the entry of budded virus into the cells (Blissard & Wenz, 1992). However, another 16 genes are specific to the group I NPVs, 3 of which also encode structural proteins (*odv-e26, ptp1, vp80a*). It is tempting to speculate that acquisition of novel structural proteins may contribute to baculovirus speciation by causing alterations in host range. Of the other group I NPV specific genes, two (*ie2* and *lef7*) are implicated in the regulation of viral gene expression and one is another *iap* (*iap1*). For all of these, it is possible to postulate an association with virus host range. The functions of the remaining group I NPV specific genes are not known.

Contrary to the group I NPVs, the group II NPVs do not share many genomic characteristics. The gene content mapping shows that the clade is supported by the loss

of one gene, the *11.3K* gene, and the acquisition of 10 genes, of which only 3 are present in all the group II NPVs genomes (figure 14). Most of these genes are of unknown function, or need to be characterised in a baculovirus context, like the *parg* (*ld141*) or *rr2* genes.

A striking feature emerging from the mapping of baculovirus genes onto the phylogeny is the number of homoplastic changes predicted. Of particular note is the number of genes that appear to have been gained independently by several lineages (figure 14). Based on these 13 genomes, 33 genes have been acquired separately at least twice and 10 of them (*ac63, ctl1, ctl2, he65, hlx2, p35, ptp2, rr1, rr2a, ld138*) three times. "Independent acquisition" here could either reflect two lineages separately acquiring a gene from the genome of another organism, or one lineage acquiring the gene which is then transferred horizontally to other baculovirus lineages. Further study will be required to determine the evolutionary history of these genes. It is important to bear in mind that these predictions should be interpreted with caution. They represent the most parsimonious interpretation of the currently available data but it is possible that, as further genomes become available, the mapping of the tree may change. Nonetheless, the emerging picture is one of baculoviruses continuously "sampling" their genomic environment (either the host genome or the genomes of coinfecting agents) for beneficial genes during the course of their evolution.

2. Discovering functional linkage

Several computational approaches have been developed to predict the function and physical interactions of proteins from genomic sequence data (Eisenberg et al., 2000, Galperin & Koonin, 2000, von Mering et al., 2002). When these approaches are applied to baculovirus genomes a number of interesting features come to light. The first is the presence of 2 putative DNA repair systems in a number of distantly related genomes. The second is the presence of host related factors in non-related viruses infecting larvae from the family Lymantriidae. The final observation is of a cluster of 4 genes, which appear to be a physically linked core baculovirus feature.

a) In silico methods to determine putative functional linkages

Protein phylogenetic profiles have been used in prokaryotes to infer protein function (Pellegrini et al., 1999). Genes that share the same phylogenetic profile are

defined as those genes that are always present in the same genomes. When they cannot be correlated with single gene acquisitions in a common ancestor, protein phylogenetic profiles can provide clues towards the understanding of protein function. The Rosetta-Stone method, which deduces protein interactions from instances of genes fusion and fission (Marcotte et al., 1999), can also be used to infer or reinforce functional hypotheses. Lastly the examination of gene neighbourhood can be useful to detect operon style gene clusters (Dandekar et al., 1998).

b) A gene of unknown function part of putative DNA repair systems

The gene mapping only reveals one example of a conserved protein phylogenetic profile that cannot be explained by the baculovirus species phylogeny. The genes ribonucleotide reductase 1 (*rr1*) and *ld138* are always present together in OpMNPV, SpltNPV, SeMNPV, LdMNPV and CpGV (figure 15A). In addition, *rr1* is itself always linked to a ribonucleotide reductase-2 gene (*rr2*). Ribonucleotide reductase is normally composed of 2 subunits encoded by 2 different genes (Reichard, 2002). In baculoviruses, there are 2 kinds of *rr2* gene, *rr2a* and *rr2b*, which are both contained within the genome of LdMNPV but otherwise have different distributions on the tree. One or both of these *rr2* genes is always found with *rr1*. A further association can be seen within the lepidopteran NPVs, where the *dUTPase* gene has a phylogenetic distribution clearly linked with the ribonucleotide reductase genes and *ld138*. It is present in OpMNPV, SpltNPV, SeMNPV and LdMNPV. The idea that these 4 genes share a functional linkage is further reinforced by the fact that the *ld138* and *dUTPase* genes are fused in OpMNPV, forming a 'rosetta stone' gene (Luque et al., 2001). Ribonucleotide reductases catalyse the production of deoxynucleotides (dNTPs) required for DNA replication and repair (Reichard, 2002). The function of the dUTPase is to prevent the mutagenic incorporation of uracyl into DNA (Vassylyev & Morikawa, 1996). These sequence homologies suggest that these genes constitute a DNA repair system in baculoviruses. Protein phylogenetic profiling and 'rosetta-stone' analysis show that *ld138* is clearly linked to this system, strongly suggesting that the protein encoded by this gene is also involved in DNA repair.

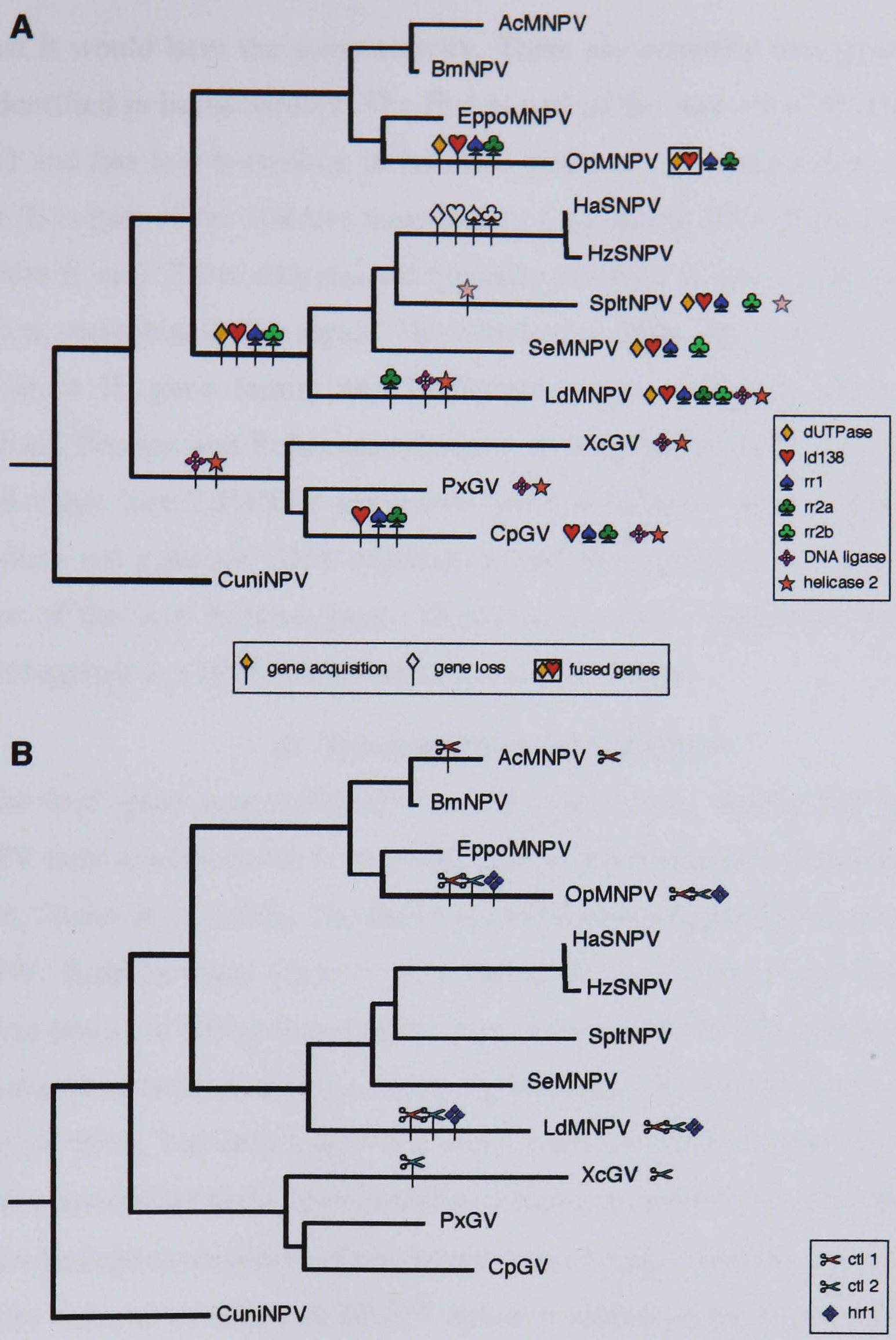


Figure 15: Phylogenetic profiles. The gains and losses of (A) dUTPase, ld138, rr1, rr2a, rr2b, DNA ligase, helicase 2, and (B) hrf1, ctl1, ctl2 are mapped onto the tree derived from the analysis of combined sequences of the 30 shared genes. The presence of these genes in individual baculovirus genomes is summarised on the right of the figure.

c) A second putative DNA repair system in baculoviruses

Another pair of genes involved in DNA metabolism has related phylogenetic profiles unrelated to baculovirus phylogeny. The DNA ligase gene is only present in genomes containing a helicase-2 gene: CpGV, PxGV, XcGV and LdMNPV (figure 15A). So far, only the SpltNPV genome appears to have a helicase-2 gene on its own. However this gene only has low homology to the other helicase-2 genes, and it is not

clear that it would have the same activity. There are currently two types of helicase genes identified in baculoviruses. The first is part of the core set of baculovirus genes (table 1) and has low homology to helicase genes of other organisms. The second, helicase-2, is part of the helicase superfamily I, including PIF1 from eukaryotes and RecD from *E. coli*. These enzymes are typically involved in DNA metabolism, such as replication, recombination or repair. The baculovirus DNA ligase is homologous to the DNA Ligase III gene family and is therefore most probably involved in DNA metabolism. Pearson and Rohrmann (Pearson & Rohrmann, 1998) have characterized the DNA ligase from LdMNPV and shown that it is an active ligase. They also showed that it does not enhance DNA replication and that helicase-2 is not a functional analogue of the core helicase gene. Therefore it is most likely that both genes are involved together in a DNA recombination or repair system.

d) Lymantrid host related factors

The *hrf1* gene from LdMNPV has been shown to expand the host range of AcMNPV both *in vitro* and *in vivo*, allowing it to infect non-permissive hosts (Chen et al., 1998, Thiem et al., 1996). The only baculovirus homologue of this gene is found in OpMNPV. Both of these viruses infect larvae of the Lymantriidae family, namely *Lymantria dispar* and *Orgyia pseudotsugata*. Two conotoxin-like genes, *ctl1* and *ctl2* are also found in both of these genomes (figure 15A). Other baculoviruses encode one or other of these, but only OpMNPV and LdMNPV encode both. The baculovirus conotoxins are similar to the ω-conotoxins, which are paralytic toxins (O'Reilly, 1997). The exact biological properties of the baculovirus *ctl* genes remains to be determined, as no effects were observed in AcMNPV deletion mutant experiments (Eldridge et al., 1992). The mode of action of these genes remains unclear but there is a clear phylogenetic link between *hrf1* and the presence of both conotoxin genes. The fact that this linkage is associated with viruses infecting a particular family of Lepidoptera suggests the proteins may function as part of a viral response to host resistance.

e) Baculoviridae core gene cluster

Examination of synteny conservation across the sequenced baculoviruses shows that gene order is poorly conserved. Even the relative arrangement of the 30 conserved genes is not maintained, showing that extensive genome rearrangements have occurred during the course of baculovirus evolution. It is thus particularly notable that a set of 4 genes, present in all baculoviruses, has remained in the same relative position in all the sequenced genomes, including the mosquito virus CuniNPV (figure 16) and newly

available baculovirus sequences (Li et al., 2002a). All the new complete baculovirus genome sequences presented in symposia this year, *Neodiprion lecontei* NPV (Arif et al., 2002), *Adoxophyes honmai* NPV (Nakai et al., 2002), *Rachiplusia ou* NPV (Harrison & Bonning, 2002), *Phthorimaea operculella* GV (Croizer et al., 2002), *Cryptophlebia leucotreta* GV (Jehle, 2002), also harbour the same arrangement for the 4 genes.

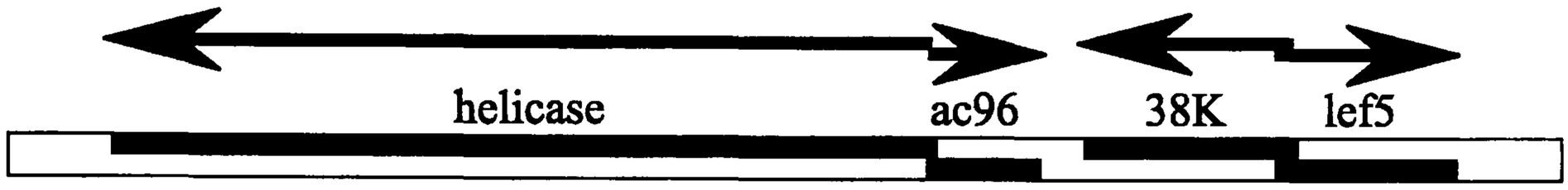


Figure 16: Schematic map of the Baculoviridae core gene cluster. The relative positions of the helicase, ac96, 38K and lef5 genes are conserved in all 13 sequenced baculovirus genomes.

Apart from this cluster of 4 genes there is not even another single pair of genes that is kept together in all the genomes. It is therefore highly unlikely that this feature is random. This cluster of 4 genes includes helicase and *lef5*, which are essential genes for baculovirus replication and transcription. It also comprises 2 genes of unknown function, *ac96* and *38K* (*ac98*) (figure 16). The fact that these genes are found in the same relative position in all genomes indicates that there is some physical constraint preventing them, or at least their DNA sequences, from being separated. One possibility is that the genes are part of a single transcriptional unit, in an operon-type arrangement or as part of a spliced transcript(s). This in turn might suggest that there is some functional linkage between all 4 genes. However the fact that they are coded by different strands of viral DNA makes a potential transcriptional/splicing link harder to envision. An alternative explanation is that the physical structure of this region of the genome is essential for virus replication and must be preserved. More detailed transcriptional mapping and functional analysis will be required to understand the significance of this striking conservation of gene order.

This conserved baculovirus cluster of 4 genes is part of a larger conserved gene cluster that has been observed among lepidopteran baculoviruses (figure 17). Gene Parity plots analyses have repeatedly outlined a number of gene clusters, the size and number of which depends on the relatedness of the viruses compared (Chen et al., 2001, Hayakawa et al., 2000a, IJkel et al., 1999, Luque et al., 2001). The lepidopteran viruses share a conserved region that extends in and around the region of the helicase gene

(figure 17). It has already been suggested that this region might be under some sort of constraint (Chen et al., 2001). However, further functional analyses need to be performed to unravel the nature of the physical constraint that links theses genes.

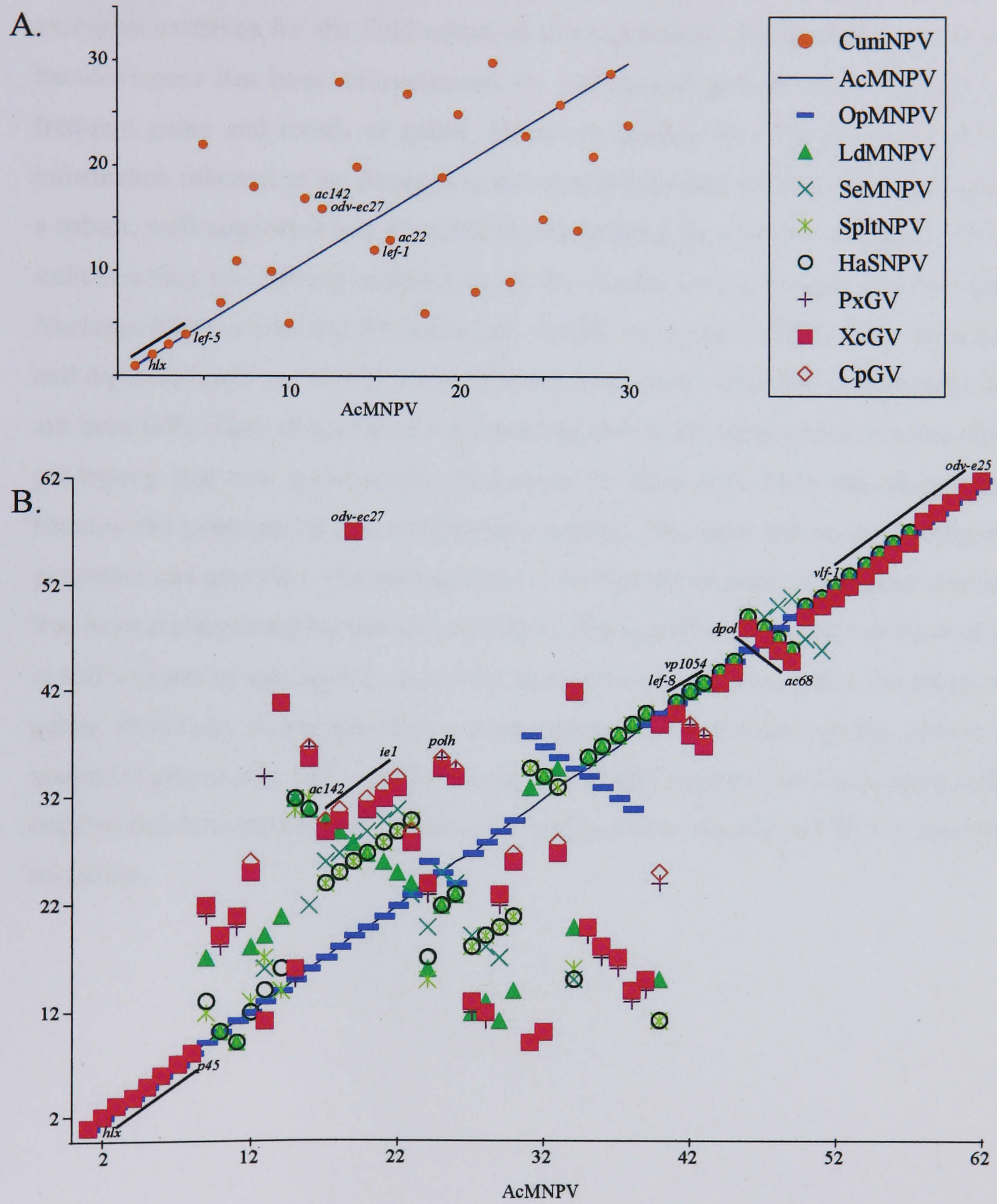


Figure 17: Gene Parity plots. A. Order of the 30 genes common to all baculoviruses. AcMNPV / CuniNPV, clusters of contiguous genes are outlined. (both small clusters are not conserved in the GVs). B. Order of the 62 genes common to all lepidopteran baculoviruses, clusters of more than 3 genes are outlined. The genes have been numbered 1-30 or 62, starting from *hlx* towards *lef-5*. AcMNPV was chosen as the X axis.

D. Summary

The comparison of complete baculovirus genome sequences has provided extensive evidence for the fluid nature of these genomes. The evolutionary history of baculoviruses has been characterised by widespread genome rearrangements and frequent gains and losses of genes. However, despite this fluidity, there is ample information inherent in the complete genome sequence data to allow the construction of a robust, well-supported and plausible tree describing baculovirus evolution. This tree indicates that the current subdivision of the family *Baculoviridae* into two genera, *Nucleopolyhedrovirus* and *Granulovirus,* should be revised. NPVs from lepidopteran and dipteran hosts are clearly more different from each other than lepidopteran NPVs are from GVs. They should not be grouped together in the same genus. The baculovirus phylogeny can now serve as the framework to place data from the comparison of baculovirus genomes into an evolutionary context. We have shown how comparative genomics can provide a fascinating view of the history of gene acquisitions and losses that have accompanied baculovirus evolution. The significance of most of these changes is still a matter of speculation, due to the lack of functional information for many of the genes. However, as our knowledge of baculovirus gene function grows, comparative genomics approaches will yield increasingly profound insights into baculovirus biology, baculovirus-host interactions, and the factors that have shaped and driven baculovirus evolution.

III. Chapter Three:

Sampling Baculovirus Diversity and Evolution

A. Baculovirus Diversity

Biological diversity can be approached from ecological and/or evolutionary perspectives. Ecologists focus on species richness at the level of local communities, and systematists on patterns of changes in diversity on a larger time scale (Futuyama, 1998). However, the basis for these investigations has to be a solid taxonomy of the organisms studied. It is crucial to be able to define and recognise what constitutes a species before undertaking any work on species diversity.

The study of baculovirus diversity has mainly been in the context of ecological studies, where virus variation has been assessed across specific isolates or host species (Cory et al., 1997). A number of cases have shown the great variation that can exist within viral populations, for example a single field isolate of *Autographa californica MNPV* containing seven distinct strains (Stiles & Himmerich, 1998) or a single infected *Panolis flammea* caterpillar reported to harbour 24 distinct NPV strains (Hodgson et al., 2001). Also a number of studies have shown geographic variation between isolates from the same hosts and local similarity of viruses from different hosts (reviewed in (Cory et al., 1997)). In short, there is much evidence for diversity of baculoviruses in the wild. However, so far these efforts have been concentrated on the differences between isolates and not on the shared similarities that would indicate the evolutionary relationships between distinct strains and isolates.

Evolutionary studies of baculovirus diversity encompass a range of baculovirus gene phylogenies (Bideshi et al., 2000, Bulach et al., 1999, Chen et al., 1999, Chen et al., 1997, Clarke et al., 1996, Cowan et al., 1994, Jin et al., 1999, Kang et al., 1998, Li et al., 2001, Liu & Maruniak, 1999, Wormleaton & Winstanley, 2001, Zanotto et al., 1993). Some efforts have been made to discuss the results of these analyses in terms of macro-evolutionary questions, such as the influence of hosts associations on baculovirus evolution (Wormleaton & Winstanley, 2001, Zanotto et al., 1993). However, these studies generally did not have a broad sampling of baculovirus diversity, either due to a lack of available sequences or to a focus on particular.

In this chapter I intend to investigate the overall phylogenetic diversity of baculoviruses. This will be an essential step towards understanding baculovirus evolution. I wish to address a number of issues including the evolutionary relationships between baculoviruses from different insect orders (Section C), the interrelationships within the lepidopteran baculoviruses and the relations between virus strains and virus

species (Section D), and lastly the influence of host use on the evolution of lepidopteran baculoviruses (Section E).

1. Baculovirus Nomenclature

Understanding baculovirus diversity is linked to taxonomy, as it is essential to identify the viruses in question in any particular study. It is quite likely that synonyms are employed to define the same virus species and that conversely, one name might refer to several species.

a) Baculovirus classification

The last meeting (June 2002) of the International Committee for the Taxonomy of Viruses led to a number of changes for the classification of viruses. The most important change is probably that each viral taxonomic determination is to be assigned a numeric code. This code becomes longer as the taxonomic level decreases from family to isolate. It allows the clear assignment of any virus isolate to a specific taxonomic description and easy cross-referencing between different strains or species (Cornelia Büchen-Osmond pers. com.).

These changes should soon affect baculovirologists who work on viral strains like AcMNPV L1, C6 or E2. When codes are eventually allocated to these most commonly studied strains, all the sequences deposited will be linked to each strain and not directly to the species AcMNPV. In the case of AcMNPV strains this does not seem to pose any problems currently, but for other baculoviruses there is some confusion existing in the genebank database. For example, the OpMNPV genome sequence is listed under the OpSNPV species. If the codes had been in place at the time of submission this would not have happened.

Table 6:Taxonomic structure of the baculovirus family*

Taxonomic level	Code	Name	Abbreviation
Family	00.006.	*Baculoviridae*	
Genus: assemblage of viral species with similar characteristics	00.006.0.01. 00.006.0.02.	*Nucleopolyhedrovirus* *Granulovirus*	NPV GV
Type Species: virus description chosen as genus representative	00.006.0.01.001. 00.006.0.02.001.	*Autographa californica multiple Nucleopolyhedrovirus* *Cydia pomonella Granulovirus*	AcMNPV CpGV
Species: assemblage of virus strains sharing similar niches	00.006.0.01.003. 00.006.0.02.002.	*Autographa californica MNPV* *Artogeia rapae granulovirus*	AcMNPV ArGV
Strain or Isolate†: viral population, genotype	00.006.0.01.003.00.000.006. 00.006.0.02.002.00.000.007.	Trichoplusia ni MNPV Pieris brassicae granulovirus	TnMNPV PbGV
Tentative species†: awaiting approval by ICTV committee	00.006.0.81.011.	Orgyia pseudotsugata SNPV	OpSNPV

*modified from ICTVdb-2002, †not in italics

Apart from the assignment of virus codes (table 6) and the additional recognition of a few virus species and isolates, the classification of the baculoviruses is mostly unchanged at this stage. One notable change is that the type species of the genus *Granulovirus* is now *Cydia pomonella GV* (CpGV) instead of *Artogeia rapae GV* (table 6).

The importance of the species concept for viruses was also emphasised by the ICTV. Researchers should attempt to identify the species of the virus they are working on. The virus strains and isolates that constitute a species can have different pathologies and host specificities. They have generally been grouped together on the basis of their genetic similarities (Drebot et al., 2002).

In the case of baculoviruses, the denomination of species is rarely used. Researchers tend to work on virus isolates and little effort is made to confirm species affiliations. A number of factors might have contributed to this. Firstly, there are actually only 36 baculovirus species listed plus eight tentative species (ICTVdb-2002). This does not encompass all the baculovirus species known to occur in the wild. For

examples, recent publications have been made on Malacosoma disstria NPV, M. californica pluviale NPV and Adoxophyes orana GV (Nielsen et al., 2002, Wormleaton & Winstanley, 2001). Based on the sequence data presented in these papers, these viruses clearly do not belong to any of the recognised species of the ICTV list and are therefore new baculovirus species. There are still a lot of baculovirus species to describe formally and to submit to the insect virus taxonomy committee. Thus, baculovirologists should be encouraged to make taxonomic statements within publications or to be more active in making ICTV submissions. Another reason for not using the ICTV classification might lie in the confusion between strains and species themselves and the subsequent conflicts that might result in the naming of baculoviruses.

b) Naming baculoviruses

Baculovirus names derive from the species name of the host they were first isolated from and from the cytopathology they cause. This describes either the morphology of the occluded virus particles or the cytopathic effect observed in infected cells. Granuloviruses have small (<0.5 μm) ovoid occlusion bodies (OB) that make the infected cells appear full of granules. The pathology they cause is called granulosis disease. Nucleopolyhedroviruses have large polyhedral occlusion bodies that form and accumulate in the infected cell nucleus. The pathology they cause is called nucleopolyhedrosis disease. For the NPVs a further distinction can be made according to the number of nucleocapsids present per virion. There can be either single (S) or multiple (M) phenotypes. Some viruses have many individually enveloped nucleocapsids per OB (SNPV; singly-embedded, single NPV) whereas others have 'bundles' of enveloped nucleocapsids within each OB (MNPV; multiply-embedded, multiple NPV) (Williams & Faulkner, 1997). This trait has no phylogenetic relevance (Zanotto et al., 1993) and its biological basis has not been determined.

The type species of the genus *Nucleopolyhedrovirus*, is named *Autographa californica multiple nucleopolyhedrovirus* (AcMNPV) because it was first isolated from the alfalfa looper, *Autographa californica*, and because its occlusion bodies are polyhedral, localised in the infected cell nucleus and contain virions with multiple nucleocapsids.

This nomenclature of the baculoviruses has the advantage of being quite simple. But there are a number of problems with this system, which are worth bearing in mind at the start of a phylogenetics project as this process can cause confusion.

(vii.) Some viruses infect more than one host species therefore there could be some synonymy whereby different names refer to the same virus.

(viii.) Some host species can be infected by more than one virus of a certain morphotype (GV, NPV, SNPV, MNPV). It is thus possible that the same name can be used for different viruses.

(ix.) Misidentification of the host might lead to an erroneous virus name.

(x.) The host name can also change after taxonomic revision. This leads to confusion because there is always a time lag between the official change and the community using the right names. This means that sometimes viruses are named after host species names that are no longer valid. A notable example is the granulovirus *Artogeia rapae GV*, which should be named *Pieris rapae GV* according to the current taxonomy of the host (*Artogeia* is a subgenus of *Pieris*, Zoological records). In principle, host synonymy should readily be found out through literature searches. However, taxonomic revisions are not always available or readily accessible for the more obscure groups.

One of the aims of this project will therefore be to try to understand what constitutes a baculovirus species and to propose species names whenever possible.

2. Baculovirus Phylogenies

Phylogenetic trees can be used at different levels of viral diversity, to characterise population structures or relationships between species (Holmes, 1999). In the case of baculoviruses, phylogenies have been used mainly to infer the relationships of particular gene sequences. The reconstructions of baculovirus trees has mostly been in conjunction with the characterisation of viral genes, such as *egt*, *chitinase, cathepsin, lef-2, gp41* (Chen et al., 1999, Clarke et al., 1996, Kang et al., 1998, Liu & Maruniak, 1999) or with the comparison of genes from a particular genomic region, such as the *egt* region of Adoxophyes orana granulovirus (Wormleaton & Winstanley, 2001).

Few baculovirus papers have a clear evolutionary focus. The early work on occlusion body proteins by Rohrmann *et al* (1981) was the first dedicated effort to understand the relationships of major baculovirus groups, sampling all morphotypes available (MNPV, SNPV, GV) and viruses from three hosts orders (Lepidoptera, Hymenoptera, Diptera). However the technique of protein N-terminal amino acid sequencing used in this study has since been replaced in phylogenetic studies by the more efficacious techniques of DNA sequencing. Zanotto *et al* (1993) published a

phylogenetic study based on polyhedrin gene sequences, which is still the reference in terms of baculovirus evolution. The analyses were based on 18 baculovirus sequences, including three GVs and one Hymenopteran baculovirus. Evolutionary rates and host associations were investigated. However the taxa sampling was not sufficient to infer such macro-evolutionary patterns with confidence. Bulach *et al* (1999) supplemented their phylogenetic analysis of the *polh* gene by sequencing fragments of the *DNA polymerase* gene for a number of lepidopteran NPVs. Their *polh* tree provided provisional indications that viruses might cluster according to the insect hosts they utilise.

One of the aims of my work is to expand on previous studies of baculovirus evolutionary history by improving the taxa sampling both in terms of virus diversity and diversity of insect hosts. Although I cannot hope to complete an exhaustive sampling of the baculoviruses, as I suspect many remain to be discovered, I aim to broaden it sufficiently to be able to observe macro-evolutionary patterns, such as historical host associations.

B. Material and Methods

This section provides a background to the samples studied and a detailed description of the molecular methods performed to obtain the DNA sequences used in the phylogenetic analyses described in the following sections (C-E). A brief outline of the methods and a detailed description of the samples used to answer specific questions is also given in each individual section.

1. The baculovirus collection

NERC-CEH-Oxford holds an extensive collection of insect viruses, which comprises about 340 baculoviruses from a wide range of hosts. The majority of the collection consists of viruses from Lepidoptera, but a number of viral samples from other orders of insects are also represented, including baculoviruses from 5 species of Hymenoptera, 2 species of Neuroptera, and 1 species of Diptera as well as non-occluded viruses from 3 species of Coleoptera. The collection contains a wide diversity of lepidopteran viruses from 80 species across 20 families of Lepidoptera. The families Noctuidae, Lymantriidae and Arctiidae, from the Noctuoidea superfamily, are the best represented with viruses from 53 moth species. Furthermore, it was possible to access sample collections of Dr Jenny Cory and some samples were generously given by a number of researchers from around the world. Appendix 2 lists the samples that have been worked on during this project, their codes and provenance.

2. Viral DNA Purification

The virus samples within the collection have been preserved in various forms and conditions and are therefore of variable quality. Some samples are clean occluded virus preparations, whereas others are infected caterpillar cadavers and many are partly purified macerates of older samples. The majority are rare and irreplaceable. This means that precautions had to be taken in order to use the minimum amount of material. Optimisation of the protocol commonly used to purify baculovirus DNA (O'Reilly et al., 1992) was therefore essential, especially as the diverse viruses used for this study have different chemical properties. For efficacy, a standardised protocol was optimised to enable DNA extraction from this diverse range of baculovirus samples.

a) Basic steps of baculovirus DNA preparation

The following points outline the basic steps that need to be undertaken to successfully purify baculovirus DNA. A range of methodologies was used in a number of combinations to determine which would be the most suitable and reliable for the purpose of these studies.

i. *Purification of the occluded viruses from insect cadavers.* The cadavers were homogenised and the occlusion bodies pelleted through a 45% sucrose cushion. If the sample remained dirty (contaminated with host material) further washes (in 0.5% SDS and in 0.5M NaCl) were performed. The size of the pellet reflected the number of occlusion bodies.

ii. *Release of the embedded virions from the occlusion bodies).* At this stage, the occlusion bodies must be dissolved. In the natural life cycle, occlusion bodies break down in the midgut of their host. Baculovirus species are quite specific to their host. Therefore, through natural selection, the occlusion bodies have been adapted to dissolve at the pH of their host midgut (pH10-10.5, (Terra & Ferreira, 1994)). This should be taken into consideration especially when dealing with non-lepidopteran species. Incubation in 0.1 M $NaCO_3$ or 0.01 M NaOH for different periods of time was tested.

iii. *DNA extraction.* The purification of DNA from occlusion-derived viruses follows the same protocols as standard purification of DNA from cellular organisms. Standard proteinase-K digestion and phenol / chloroform extraction, followed by ethanol precipitation or dialysis were tested, as well as column based DNA purification kits.

b) Standard microfuge protocol for the purification of Baculovirus DNA

The following sections describe the extended protocol that was routinely used. This protocol was found to be the most efficient and the least prone to possible sample cross-contamination, as it requires only a minimum amount of pipetting and manual sample transfers from tube to tube. Depending on the sample condition the DNA purification was started at point i, ii, or iii. The whole protocol was performed in standard 1.5 ml eppendorf tubes using a standard benchtop centrifuge (such as Eppendorf 5415C)

i. *Recovery of occlusion bodies from insect cadavers.* A volume equivalent to 200µl of insect cadavers was taken. The mixture was crushed with a blue tip heat-moulded

to fit the bottom of an eppendorf tube. 250µl of H_2O was added and the sample was homogenised by vortexing prior to centrifugation at 8000 rpm (5000g) for 10 min. The supernatant was discarded and the pellet resuspended in 100µl H_2O.

ii. *Occlusion body clean-up.* A succession of washes was undertaken on the samples to obtain occlusion bodies, which were as clean as possible. For the first wash, 250µl of H_2O was added to a 100µl sample (from above or starting at this stage) and the sample was homogenised by vortexing prior to centrifugation at 8000 rpm (5000g) for 10 min. The supernatant was discarded and the pellet resuspended in 250µl 0.5% SDS and vortexed prior to centrifugation at 12000 rpm (7500g) for 5 min. The supernatant was discarded and the pellet resuspended in 250µl 0.5M NaCl and vortexed prior to centrifugation at 12000 rpm (7500g) for 5 min. The supernatant was discarded and the pellet resuspended in 50µl H_2O.

iii. *Occlusion body dissolution.* The crystalline protein stucture of the occlusion bodies was broken down in strong alkaline conditions. This treatment, modified from (Moser et al., 2001), was the strongest tried and should therefore work for all the samples studied. To the above 50µl sample, 5µl of 0.1M NaOH (pH 12.5) was added. The incubation was performed at room temperature for 15 min. the sample was then neutralised with 5µl of 100mM Tris-HCl pH 8.

iv. *DNA purification.* The DNeasy kit (Qiagen) was used to proceed with the DNA extraction, with no modification from the manufacturer's standard protocol for animal tissues. The precise composition of the buffers supplied with the kit was not provided. To the 60µl viral sample, 180µl of buffer ATL (digestion buffer) and 20µl of proteinase-K were added. The digest was performed overnight at 37°C. The buffer of the lysate was then adjusted (buffer AL and ethanol) to provide optimum DNA binding to the column membrane. The first centrifugation of the columns removed most of the contaminants from the DNA that is selectively bound to the membrane. Further washes were performed (ethanol based buffers AW1 and AW2) to remove the remaining contaminants. The DNA was then eluted with 200µl of 10mM Tris (pH 8.5) buffer.

v. *DNA quantification.* The quantity and quality of the DNA were evaluated by running 5µl of the DNA preparation on 1% agarose-TBE gel (Sambrook et al., 1989). This gives a crude estimate of the DNA concentration and will not detect DNA of concentration less than 2 ng/µl. This method was preferred to the more sensitive detection method by spectrophotometer, because of the smaller amount of

material lost in the process. The smallest volume required for an OD measurement is 100µl in non-disposable cuvettes. Cross-contamination between samples would be quite likely in this process due to the use of a single cuvette, even with thorough washes between measurements. So the samples would have had to be discarded after quantification. Because baculoviruses genomes are fairly small and, of low complexity, 100pg should be sufficient for efficient PCR amplification. Therefore, samples that yield no detectable DNA by gel electrophoresis and ethidium bromide staining might still contain enough DNA to produce good PCR products.

3. Primer design

Table 7: Number of baculovirus sequences available in the databases early in 2000.

Gene	Baculovirus sequences	Phylogenies (non-exhaustive list)
polyhedrin	46	Zanotto et al (1993), Bulach et al (1999)
DNA polymerase	14	Bulach et al (1999)
egt	19	Clarke et al (1996)
gp37	11	NA
lef-2	17	NA
odv-e56	10	NA
vp1054	6	NA
Complete genomes	6	NA

Using sequences available in the public domain (GeneBank), gene alignments were performed in order to define regions sufficiently conserved across a wide range of baculoviruses to design oligonucleotides. The genes initially investigated for primer design included *polh*, *dnapol*, *lef2*, *egt*, *odv-e56*, *gp37*, *vp1054* based on the number of baculovirus species that had been sequenced for those genes and taking into consideration which genes had previously been used for phylogenetic studies (table 7). For the purpose of complementing existing phylogenies, the *DNA polymerase* and *polyhedrin* genes were the most obvious candidate genes. Primers were designed and tested for these 2 genes (table 8). However, the *dnapol* primer sets did not amplify from a wide array of baculovirus samples.

The comparison of 9 complete genome sequences and their study in an evolutionary framework highlighted the genes that were most suitable for phylogenetic studies (Herniou et al., 2001). Only a few of the genes that were found to be present in all the complete genomes, gave the same tree as the combined gene sequences alignment (table 3, chapter 2). These genes include *ac22, ac81, ac119, ac142, ac145, lef-8, lef-9.* The completion of additional genome sequences shortened that list to 6 genes, with the removal of *ac145* from the lists of genes common to all baculoviruses. Out of these 6 genes, *lef-8* (RNA polymerase subunit) and *ac22* (gene of unknown function) were picked as suitable candidates for the phylogenetic study of the diversity and evolution of baculoviruses, because they contained conserved protein motifs at suitable intervals to design oligonucleotides (table 8). The *ac22* gene was also picked because a sequence of this gene was available for a hymenopteran baculovirus, *Neodiprion sertifer NPV.*

Table 8: Design of degenerate oligonucleotide primers used for the amplification of baculoviruses

Gene		Oligonucleotide sequences §	Amino acid motif †	Size*
polh	PF	gtaaaacgacggccagtTYIKIGGICCIGGIAARAA	F(LR)GPGKN	600
	PR	aacagctatgaccatgTCIGGIGCRAAYTCYTT	KEFAP(DE)	
	PFHYM	gtaaaacgacggccagtCCNMGNACNGGNCCNGGNAA	PRTGPGK	
lef-8	L8F1	gtaaaacgacggccagTTYTTYCAYGGNGARATGAC	FFHGEMT	800
	L8R1	aacagctatgaccatGGNAYRTANGGRTCYTCNGC	TAEDPY(IV)P	
	L8F2	gtaaaacgacggccagtNNNACNRCNGARGAYCC	XTAEDP	
	L8R2	aacagctatgaccatgMMNCCYTTYTGNCCRTG	HGQKGV	450
ac22	Ac22F	gtaaaacgacggccagtGGWNNTGYATNSGNGARGAYCC	W(TSN)CI(AP)EDP	400
	Ac22R	aacagctatgaccatgRTYNCCRCANTCRCANRMNCC	G(EVF)C(ED)CG(DN)	
dpol	1F	gtaaaacgacggccagtAYRYIAAYMGIGTICAIATGC	DVNRVHMQ	950
	4R	aacagctatgaccatgYTRTTRAAIAWRTAIGGRTC	DPYFFNK	1200
	5R	aacagctatgaccatgGIKRIARIACYWTICCICC	GGKVLSP	
	4F	gtaaaacgacggccagtSIGAYCCITAYWTITTYAA	DPYFFNK	800
	6R	aacagctatgaccatgCATIAIRTTYTCIARIKCCAT	MALENVM	

§ M13 primers sequences are in lowercase, degenerate baculovirus primers are in uppercase, I is for inosine; † amino acid sequence corresponding to the primer site (single letter code) *expected size (bp) of the amplification product

Although phylogenetic studies have shown that the *polh* gene might not be the best suited gene to reconstruct the phylogeny of baculoviruses (Clarke et al., 1996, Herniou et al., 2001), its importance in terms of the number of sequences available, the number of published phylogenies and its role in baculovirus biology, supported its choice as a third gene to use.

Amplification from a wide range of baculoviruses species meant that degenerate primers had to be made. However, direct sequencing is not possible with highly degenerate primers. To avoid an additional cloning step, M13(-20) and M13R primer sequences were added at the 5' end of the primers in order to be able to directly sequence the PCR product using these universal primers (figure 18). Direct sequencing of PCR products has the advantage of being fast and also producing sequences that reflect all of the PCR products and hopefully all of the genotypes present in the original sample. This in turn is a double-edged sword. If the viral sample comprises a genotypically homogeneous population, the sequence is going to have a strong clean signal. Conversely, if the sample is very heterogeneous, the sequences are going to show multiple peaks that could be impossible to interpret. Cloning would circumvent this difficulty, but create more confusion in the mixing and matching of sequences from different genes, resulting in a series of sequences from different genomes that cannot be correlated between genes.

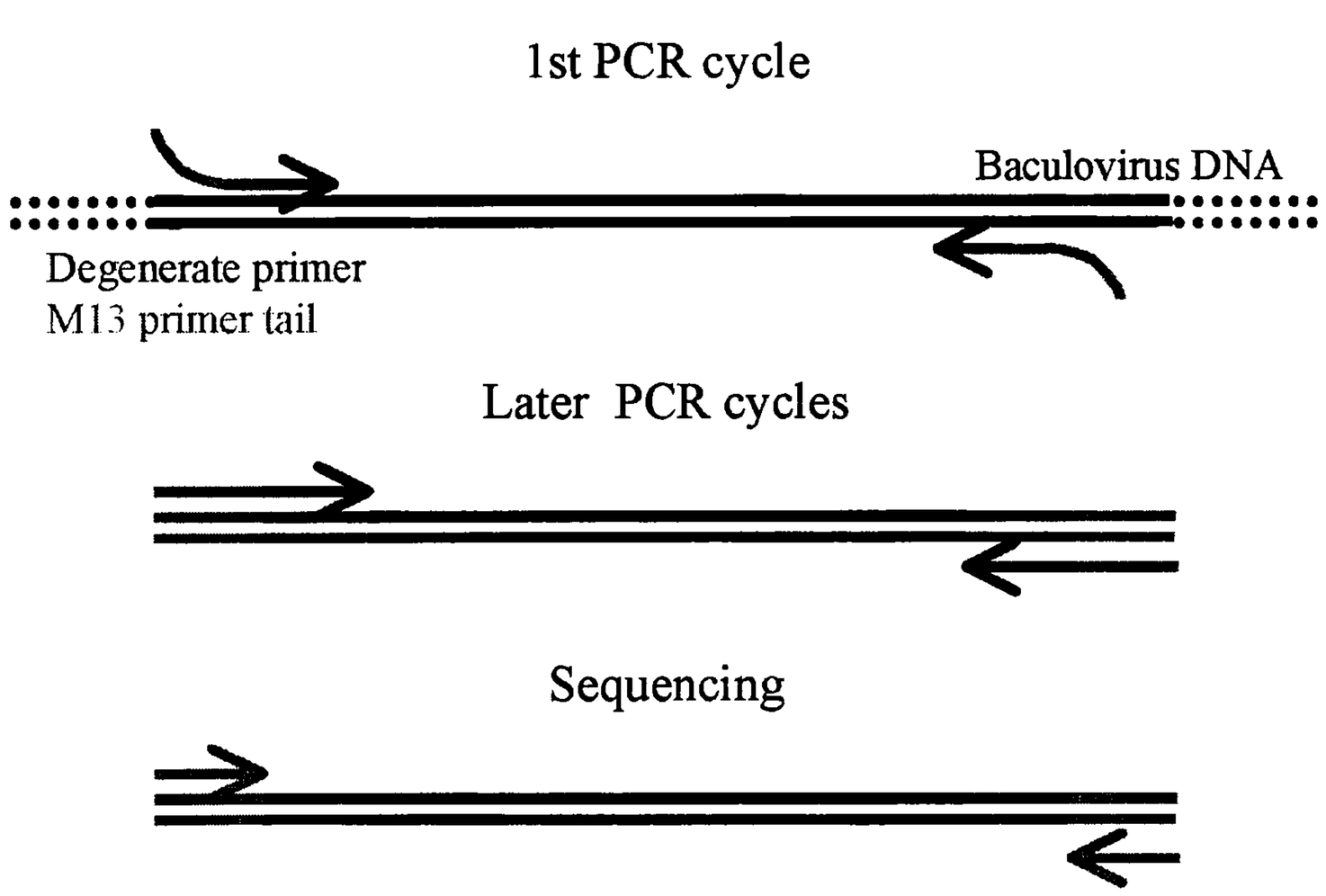


Figure 18: Degenerate primer design and strategy for the PCR and direct sequencing of baculovirus genes.

4. PCR amplification and sequencing

PCR amplifications were performed using Ready-to-Go PCR beads (Amersham/Pharmacia) as per the manufacturers instructions. The beads contain the PCR buffer, the *Taq* polymerase and the dNTPs in a lyophilised condition. To set up the PCR reactions, only template, primers and water need to be added to the tube. For 25µl

PCR reactions, 10-100 ng of viral DNA (1-2.5µl) and 25-50 pmol of primers were used.. Amplification cycles were as follows: (95°C 5min + (94°C 60s, 42°C 45s, 72°C 60s) x10 + (94°C 45s, 60°C 30s, 72°C 60s) x25 + 72°C 5 min) or a touch down cycle: (95°C 5min + (94°C 30s, 55 to 43°C 20s, 72°C 30s) x3 down 3°C after 3 cycles; 18 cycles + (94°C 30s, 60°C 20s, 72°C 45s) x20 + 72°C 5 min).

PCR products (1µl) were visualised on a 1% agarose-TBE gel (Sambrook et al., 1989). Successful amplifications were then purified using the Qiaquick PCR purification kit (Qiagen). The PCR products were bound to a membrane through a column. Contaminants, including DNA fragments smaller than 100bp were removed. The PCR products were eluted in 30µl of 10mM Tris (pH 8.5) buffer.

Cycle sequencing of the PCR products was performed for each strand with the M13(-20) and M13R universal primers using the Big Dye terminators reaction mix (Applied Biosystems). The 10µl reaction contained: 2µl of ready mix, 2µl of sequencing buffer (200mM Tris, 5mM $MgCl_2$, pH9), 0.5µl of 3.2µM primer, 2-4 µl PCR product, completed with H_2O. The sequencing cycle was: (96°C 5 min (96°C 30s, 50°C 30s, 60°C 2min) x25). Afterwards, the reactions were ethanol precipitated as per recommendations from the manufacturer (ABI). The sequences were run on a 3700 ABI automated sequencer.

Chromatograms were checked and contiguous sequences of the forward and reverse reactions assembled in Sequencher™ 4.1 (Gene Codes corporation). BLAST searches (Altschul et al., 1990) were performed on all the new sequences to verify their authenticity. The PCR and sequencing reactions were repeated at least once when the sequencing signal was not clear.

5. Phylogenetic analyses

a) Alignments

The DNA sequences obtained from the PCR fragments were aligned against sequences of the same genes available from the databases. The alignments were trimmed to the size of the PCR fragments. They were performed in MacClade 4 (Maddison & Maddison, 2000) using the option of colouring DNA triplets according to the amino acid they code. This option made aligning most DNA sequences by eye extremely easy. In cases where the aligned sequences were more divergent, BLAST search results were used as a guide.

The alignments were also translated into amino acids, allowing the reconstruction of phylogenies from protein as well as DNA sequences.

b) Phylogenetic reconstruction

All phylogenetic analyses were performed in PAUP* (Swofford, 2001.). The DNA sequence alignments were preferred to the protein alignments as they allow a larger flexibility in the choice of algorithms and methods to use. The protein alignments were analysed only as a comparison to their DNA counterparts.

Maximum likelihood (ML) was the method of choice for this study. ML methodology comprises reconstructing phylogenies based on specified models of character substitutions. The alignments were analysed using a statistical model-fitting approach implemented in MODELTEST 3.06 (Posada & Crandall, 1998, Posada & Crandall, 2001). This program selects the model of DNA substitution that best fits the data by calculating the likelihood of distance trees derived using the different models. It uses models of increasing complexity, implementing more and more parameters, such as models of substitution rate among nucleotides (like GTR, HKY85, TrN), percentage of invariable sites (I) and variation of substitution rates across sites (G). The selected model is then used to calculate a tree using neighbor-joining (NJ), minimum evolution (ME) method, or using a heuristic approach (ML) involving some branch swapping of the NJ tree by nearest neighbour interchange (NNI) to find shorter trees. The extensive computing time required does not allow for extensive branch swapping such as performed by TBR (Tree bisection reconnection) branch swapping under ML parameters. In combination with ML models, NNI branch swapping has been found to perform well compared to other search methods (Takahashi & Nei, 2000).

The robustness of the tree topologies were evaluated by bootstrap analysis under the following conditions: distance method using ML parameters as selected before with 1000 replicates, uncorrected p distance 1000 replicates using the BioNJ tree building method (Gascuel, 1997) and maximum parsimony (MP) methods with 1000 replicates. For the MP reconstructions, uninformative characters were excluded from the data matrices, the trees were built by stepwise addition and TBR branch swapping performed to find the best MP tree at each replication step.

The analyses were also made using the genome tree (30-gene tree, figure 12, chapter 2) as a backbone constraint for the tree building reconstruction. This produces trees where the basic relationships of the 13 genomes are kept as in the genome tree and where the relationships of the other viruses are reconstructed around this basic topology.

The constrained trees were compared to the unconstrained trees using one-tailed Kishino-Hasegawa (KH) and Shimodaira-Hasegawa (SH) tests implemented for ML tree scores in PAUP*(Swofford, 2001.). Any differences in tree topology could be due to errors in tree reconstruction due to lack of data or could reflect that the gene used for the analyses has a different evolutionary history to the virus genomes.

6. Looking for extra genetic information

The sequencing results obtained from non-lepidopteran baculoviruses highlighted the need for additional sequences. However, designing pairs of primers that would work on these viruses proved difficult. Often the genes common to all baculoviruses will have few conserved motifs. Finding two 6 amino acid long motifs (18 bp) at 125-350 AA interval (375-1050 bp) is not always possible. In fact for a number of genes, only one motif suitable for primer design was found conserved across all baculoviruses. This observation prompted the design of a 1 primer PCR protocol. If such a system worked, it would be more direct and less costly than a shot-gun sequencing project, which determines large portions genomic DNA sequences, but at random.

a) Key steps for a 1 primer PCR

PCR requires two oligonucleotide primers for exponential DNA amplification. Therefore the way around the limitation of only having one appropriate sequence for primer design, is to link a known sequence to the DNA of interest, and perform the PCR with a primer based on this known DNA and with the second primer targeting our gene of interest. Cutting the genomes of interest with restriction enzymes enables the ligation of DNA adaptors. With these basic steps in mind, I looked for a protocol or a kit that could be modified for the purpose. The TOPO Walker kit (Invitrogen) seemed to provide the right package with enough flexibility.

b) TOPO Walker kit aims and principles

This kit is specifically designed to enable the amplification and sequencing of unknown DNA through the use of adjacent DNA of known sequence. It uses *Vaccinia* virus topoisomerase I to join a linker that contains primer binding sites to the unknown end of a desired DNA fragment, allowing subsequent PCR amplification (figure 19). It is designed for the completion of shot-gun sequencing projects and was developed for samples such as bacterial, yeast, or phagemid artificial chromosomes.

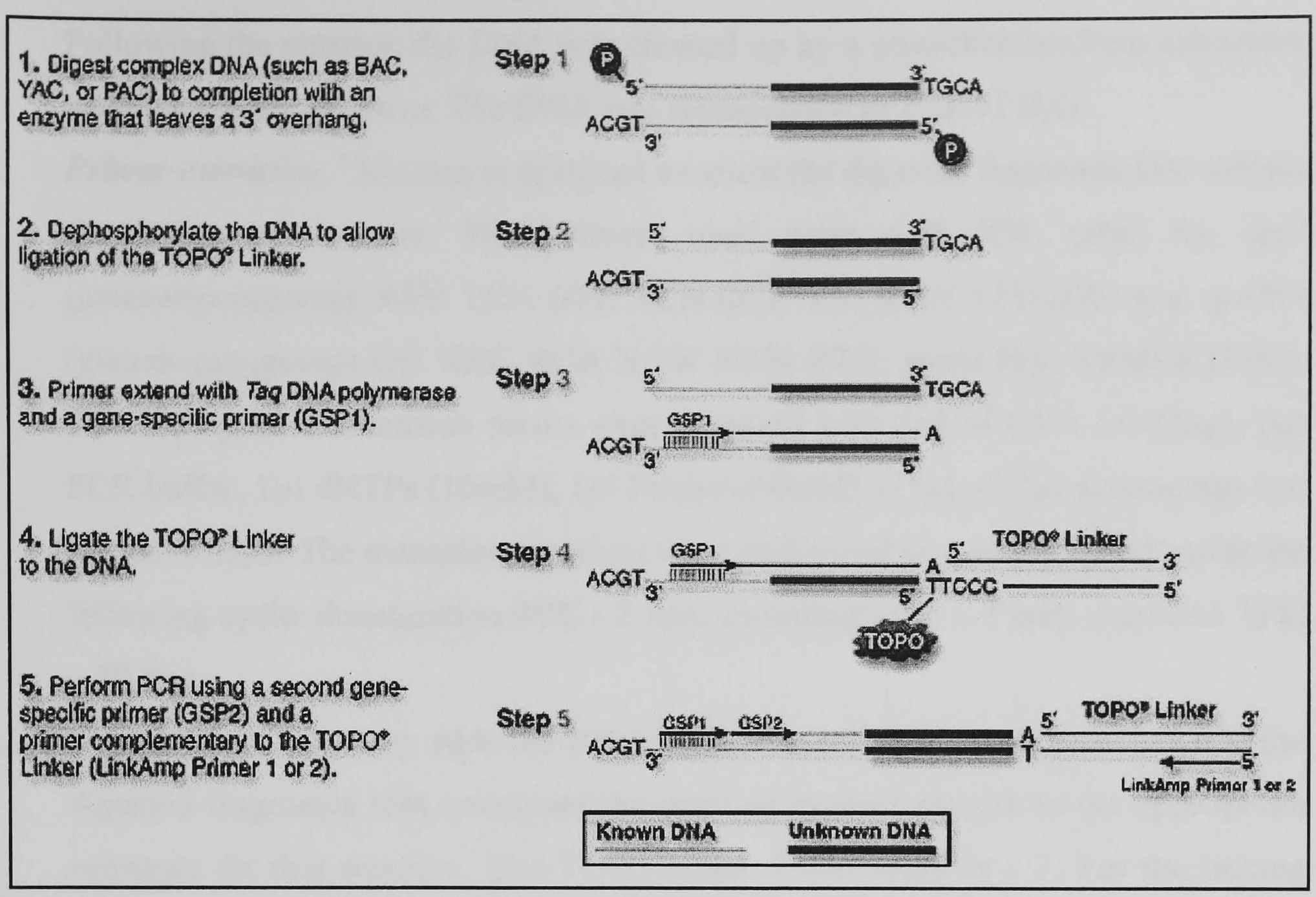


Figure 19: Schematic description of the Topo Walker kit protocol (Invitrogen).

c) Protocol modified for the amplification of baculovirus fragments

i. *Baculovirus DNA digest.* A selection of restriction enzymes was tested. The kit requires enzymes that leave a 5' overhang. Ideally the enzyme should cut the DNA into fragments of 1kb but not cut in the primer site. *PstI*, *BanII* and *BsrI* (New England Biolabs, Neb) were tested using their specific conditions and buffers. *BsrI* activity is optimum using Neb buffer 3 (100mM NaCl, 50mM Tris-HCl, 10mM $MgCl_2$ 1mM dithiothreitol, pH 7.9) at 65°C. For *PstI* and *BanII* optimum activity is achieved at 37°C with Neb buffer 3 and 4 (50mM Potassium Acetate, 20mM Tris Acetate, 10 Magnesium Acetate, 1mM dithiothreitol, pH 7.9) respectively. The protocol required starting with 500ng of DNA. The 40µl digestion reactions were prepared with 25µl of baculovirus DNA (~20ng/µl), 4µl of 10x Neb buffer, 4µl of restriction enzyme and 7µl of H_2O. The reactions were left to incubate at 37°C or 65°C for 2 hours.

ii. *Dephosphorylation reaction.* The digested DNA has to be dephosphorylated to create a substrate for the topoisomerase. To the 40µl digestion reaction, 50µl of H_2O, 6µl of 10x dephosphorylation buffer and 4µl of calf alkaline phosphatase were added. The incubation for the dephosphorylation was carried out at 37°C for 1 hour.

Following the reaction the DNA was cleaned up by a phenol/chloroform extraction and ethanol precipitation. The DNA was resuspended in 20µl of H_2O.

iii. *Primer extension.* This step is designed to select the digested fragments that contain the genes of interest. The primers used were *polh* (PR, table 8), *dpol* (gtaaaacgacggccagt SWR TCN GTR TCN CCR TA, motif YPDTDS) and *vp1054* (gtaaaacgacggccagt GG NRK RCA NAW NAN RTG, motif H(ILV)(MF)C(PI)P). The 20µl primer extension mixes were prepared with 5µl of DNA (>100ng), 2µl PCR buffer, 1µl dNTPs (10mM), 1µl Primer (50mM), 0.5µl of *Taq* polymerase and 10.5µl of H_2O. The extension reactions were performed in a thermocycler under the following cycle: denaturation 94°C - 5 min, annealing 45°C – 1 min, extension 72°C – 20 min.

iv. *TopoLinking.* This step adds the linker to the extended DNA fragments. Only the digested fragments that contained the gene of interest should be the appropriate substrate for this reaction. The TOPO-linker 3'-overhang is a T. For the linking reaction to work the DNA fragments needed to have an A in their 3'-overhang. This was achieved in the primer extension reaction as *Taq* polymerase naturally adds an A at the end of the extended PCR products. To set up the TOPO-linking reactions, 1µl of TOPO Linker and 1µl of PCR buffer were added to 8µl of primer extension product. The reaction was incubated at 37°C for 5 min.

v. *PCR.* Finally the DNA was ready to use for a PCR reaction. The kit recommended the use of a primer nested within the first gene specific primer in combination with the primer binding the linker. This was not possible in our case. Therefore the same gene specific primer as for the extension reaction was re-used for the PCR reactions. The reactions were performed using Ready-To-Go PCR beads (Amersham). The PCR reaction mixes included 2µl of Topo linking product, 1µl of each primer, 21µl H_2O added to the PCR beads. The PCR cycles were as follow: 95°C – 5 min, touch down (94°C – 15s, 55>45 (-1) – 15s, 72°C – 15s) x15, (94°C – 20s, 45– 15s, 72°C – 20s) x20, 72°C – 5 min. The PCR products were then checked on a gel. If a single band was detected, sequencing reactions were performed directly on these products. Otherwise a second round of amplification was performed with a nested primer of the linker.

~~3.~~C. Position of non-lepidopteran baculoviruses within the *Baculoviridae*

In this section, the steps that were undertaken to understand the deeper level of baculovirus diversity are described. The aim of this section was to try to answer the following question. How do baculoviruses isolated from non-lepidopteran hosts relate to each other and to lepidopteran baculoviruses? This is a fundamental question that has yet to be answered. Although previous studies have shown that baculoviruses isolated from Hymenoptera or Diptera are distantly related to those of Lepidoptera (Moser et al., 2001, Zanotto et al., 1993), there were no phylogenetic studies clustering sequences of baculoviruses from the 3 insects orders. Furthermore, there were no sequences available for baculoviruses from other arthropod orders. Any molecular data that could be gathered from such viruses would be extremely informative.

A number of issues needed to be addressed for this purpose. The foremost concerned the labelling of the samples stored in the Oxford collection. For most samples, little information was available to verify the accurate identification of the viruses, or the host taxonomic identification. Most of the samples had been collected more than 30 years ago. It was therefore impossible to go back to the original description and collection data. Some of the viruses present in the collection are unique. Their uniqueness however could reflect a number of problems. The host and/or the virus could have been wrongly identified, making the sample useless. Although it would be extremely exciting if all the samples present in the collection were the genuine article, caution should be used when dealing with 'viruses' that have never been re-isolated in the field and that have not been fully described.

1. Samples examined

A total of 18 virus samples from 4 orders of insects were examined in this study (table 9). These include definite baculoviruses such as the *Neodiprion sertifer* NPV samples, some viruses from other virus families such as the *Tipula paludosa* IV (*Iridoviridae*), but also some far less characterised samples such as the *Chrysopa* PV, which could refer to a baculovirus or a cypovirus (*Reoviridae*). The samples were

collected in the UK, apart from the *Oryctes rhinoceros* virus, which came from the Seychelles and *Neodiprion lecontei* NPV, which came from Canada.

Table 9: Viruses from non-lepidopteran insects examined in this study

Insecta order	Host species	Phenotype§	Code	Accession*
Coleoptera	*Oryctes rhinoceros*	NOV		576
Diptera	*Tipula oleracea*	NPV	TiolNPV	35
(craneflies)		NPV		281
	Tipula paludosa	IV		419
Hymenoptera	*Anoplonyx destructor*	NPV		177
(sawflies)	*Gilpinia hercyniae*	NPV	GiheNPV	i7
		NPV		K14
		NPV		239
	Nematus olfasciens	NPV		76
	Neodiprion lecontei	NPV	NeleNPV	726
	Neodiprion sertifer	NPV	NeseNPV	345
		NPV		411
		NPV		413
		NPV		K5
		NPV	Virox	M1
Neuroptera	*Chrysopa sp*	PV		330
(lacewings)	*Hemerobius sp*	NPV	HemeNPV	318
		PV		404

§ NOV: non-occluded virus, NPV: nucleopolyhedrovirus, IV: iridescent virus, PV: polyhedrovirus. *numeric codes refer to samples from the NERC-CEH Oxford virus collection, codes starting by a letter refer to samples from other sources.

DNA was purified from all samples as described above. High molecular weight DNA could be detected on agarose gels (10 – 250 ng/µl) for most samples. No detectable DNA was obtained from the isolates of the *Oryctes rhinoceros* NOV-576, the sawflies viruses *Anoplonyx destructor* NPV-177, *Gilpinia hercyniae* NPV-i7, *Nematus olfasciens* NPV-76, and *Neodiprion sertifer* NPV-M1, from the bioinsecticide Virox, or from the lacewings viruses *Chrysopa sp* NPV-330 and *Hemerobius sp* NPV-318 &-404. Nonetheless all DNA extractions including those that contained little or no DNA were used for PCR.

2. PCR & Sequencing results

For this study 3 genes were investigated: *polyhedrin*, *lef-8* and *ac22*. For *polh* 2 pairs of primers were used: a general pair PF+PR and, at a later time, another forward

primer (PFHYM) more specifically designed against sequences from hymenopteran viruses based on the amino acid sequences of *Neodiprion sertifer* NPV (Zanotto et al., 1993), *N. abietis* NPV (genebank accession AAM95580) and *N. lecontei* NPV (sequence kindly given by Basil Arif and Hilary Lauzon). For *lef-8* it soon became clear that the general primer pair L8F1+L8R1 might not work to amplify all *lef-8* homologues, as the mosquito virus CuniNPV had a different protein motif at the forward primer site. A second pair of primers was therefore designed to accommodate the widest baculovirus diversity: L8F2+L8R2. This pair of primers should amplify a fragment that is contiguous to but non-overlapping with the first one. The *ac22* primer pair was designed from a set of baculovirus sequences that included CuniNPV and a sequence from *Neodiprion sertifer NPV* (genebank accession AF121349).

a) polyhedrin

At first *polh* seemed to give the best results, as most samples produced PCR amplification products. However, blast searches showed that most sequences had extremely high similarity (scores>300) to lepidopteran baculovirus sequences (table 10). Further sequence alignments showed that these sequences were almost identical to PhopGV, PbGV, or LdMNPV. This was unexpected as these viruses are thought to be very divergent from the other sequences in the database. Thus a doubt was thrown on their authenticity and many efforts were made to reproduce the results taking great care to avoid contamination. Re-amplification of the *polh* fragments from the same samples confirmed most of the sequence results. However when the DNA was re-extracted from the original isolates, PCR amplifications gave different results, when they worked at all. It is quite probable that all these results reflect some sort of contamination, either at PCR level or at the level of the original sample. None of the sequences were retained for phylogenetic analyses.

The most extraordinary results came from 2 isolates, which based on their description should not possess a *polh* gene. The *Oryctes* rhinoceros beetle virus PCR amplification gave 2 products: one was similar to a bacterial ABC transporter ATP-binding protein and the other was polyhedrin (table 10). Both sequences are surprising in their own way. The non-occluded beetle virus was expected not to contain the major occlusion body protein gene. The second surprise was that the PCR conditions also amplified a gene that has nothing to do with *polh*. Whether either of these sequences genuinely comes from the genome of the *Oryctes rhinoceros* virus remains to be determined.

Table 10: Results of the similarity searches for the non-lepidopteran viral sequences

Primer pairs	Order	Species	Sample	Best match: BlastX	Score[§]	e-value
ac22F+ac22R	Hymenoptera	NeseNPV	345	NeseNPV Q9QAB4 ORF22	250	6.E-66
ac22F+ac22R	Hymenoptera	NeleNPV	726	NeseNPV Q9QAB4 ORF22	156	1.E-37
ac22F+ac22R	Hymenoptera	Gilpinia NPV	i7	NeseNPV Q9QAB4 ORF22	127	5.E-29
L8F2+L8R2	Neuroptera	Hemerobius NPV	404	LdMNPV Q9YMS3 lef- 8.	**303**	**1.E-81**
L8F2+L8R2	Hymenoptera	NeseNPV	413	HaSNPV Q99H09 Lef-8	85	5.E-16
L8F2+L8R2	Hymenoptera	NeseNPV	345	HaSNPV Q99H09 Lef-8	85	5.E-16
L8F2+L8R2	Hymenoptera	NeleNPV	726	HaSNPV Q99H09 Lef-8.	85	5.E-16
L8F2+L8R2	Hymenoptera	Gilpinia NPV	i7	HaSNPV Q99H09 Lef-8.	76	4.E-13
L8F2+L8R2	Diptera	Tipula NPV	35	Hz-1 virus NC_004156 Orf55	53	2.E-06
PF+PR	Diptera	Tipula NPV	35	PhopGV AAM70199 granulin	**378**	**1.E-104**
PF+PR	Hymenoptera	NeseNPV	413	PhopGV AAM70199 granulin	**371**	**1.E-102**
PF+PR	Hymenoptera	Gilpinia NPV	i7	SpltNPV AAL01689 Polyhedrin	**370**	**1.E-101**
PF+PR	Hymenoptera	Gilpinia NPV	i7	PbGV P06502 Granulin	**365**	**1.E-100**
PF+PR	Hymenoptera	Nematus NPV	76	BusuNPV P33732 Polyhedrin	**348**	**3.E-95**
PF+PR	Neuroptera	Hemerobius NPV	404	LdMNPV Q8QQA1 Polyhedrin	**338**	**4.E-92**
PF+PR	Hymenoptera	NeseNPV	345	PbGV P06502 Granulin	**337**	**7.E-92**
PF+PR	Hymenoptera	NeseNPV	413	PbGV P06502 Granulin	**332**	**3.E-90**
PF+PR	Neuroptera	Hemerobius NPV	404	PhopGV AAM70199 granulin	**324**	**8.E-88**
PF+PR	Coleoptera	Oryctes NOV	576	SeMNPV NC_002169polyhedrin	148	2.E-35
PF+PR	Diptera	Tipula IV	419	Chilo iridescent virus Q91FB0 415R.	122	4.E-27
PF+PR	Neuroptera	Chrysopa NPV	330	LdMNPV Q8QQA1 Polyhedrin	118	6.E-26
PF+PR	Coleoptera	Oryctes NOV	576	Staphylococus aureus NC_002758 ABC transporter ATP-binding protein	68	2.E-13
PF+PR	Coleoptera	Oryctes NOV	576	(168-356) Oceanobacillus iheyensis 14056 ABC transporter ATP-binding protein	75	8.E-13
PF+PR	Coleoptera	Oryctes NOV	576	(2-94) MacoNPVB AAM94988 Polyhedrin	68	8.E-11
PF-HYM+PR	Hymenoptera	NeleNPV	726	NeabNPV AAM95580 Polyhedrin	82	6.E-15
PF-HYM+PR	Hymenoptera	NeseNPV	411	NeabNPV AAM95580 Polyhedrin	80	2.E-14
PF-HYM+PR	Hymenoptera	NeseNPV	345	NeabNPV AAM95580 Polyhedrin	68	8.E-11
PF-HYM+PR	Hymenoptera	Gilpinia NPV	K14	NeabNPV AAM95580 Polyhedrin	49	9.E-06
PF-HYM+PR	Neuroptera	Hemerobius NPV	404	HcNPV Q9DY25 Polyhedrin	47	2.E-04

[§] BLAST scores higher than 300 are outlined in **bold**. These are highly significant similarity and are therefore suspicious, considering the nature of the viruses the sequences are supposed to be derived from.

The second sample that was not expected to give any result at all is the Iridovirus from *Tipula paludosa*. The pair of primers designed against *polh* amplified a sequence homologous to the Chilo iridescent virus 415R gene (table 10). The 415R gene has no homology with any baculovirus gene. Although the reason for the PCR amplification of this sequence is obscure, it appears to be a genuine sequence derived from the genome of an iridescent virus, as indicated in this isolate description. This sequence cannot be used in the phylogenetic analyses, as it has no homology to any baculovirus genes.

However, it will be deposited into genebank as an interesting accidental by-product of PCR.

The lack of convincing results from the PF and PR primers prompted a re-evaluation of the primers used for the amplification of polyhedrin sequences from non-lepidopteran baculoviruses. The sequence of the *polh* gene from the baculovirus of the sawfly *Neodiprion abietis* (NeabNPV) became available in July 2002. In addition, sequences from another sawfly virus from *N. lecontei* (NeleNPV) were kindly shared by Hilary Lauzon and Basil Arif. These 2 sequences were fairly divergent to the *polh* sequences from the GVs and lepidopteran NPVs. They became the model on which to base a new sawfly virus specific primer (PF-HYM). However, PCR amplification using this primer with the PR primer gave the strangest of results. Similarity search results showed that the sequences amplified were most similar to *polh* of NeabNPV (table 10), as expected. But the scores were low and there was little sequence homology with the *polh* of the database even at amino acid level. Aligning these sequences to any other *polh* sequences proved difficult and the many stop codons present in the sequences rendered the alignments absolutely unconvincing. Although these sequences were most similar to a *polh* sequence in the database, it is clear in my mind that they do not represent genuine *polh* sequences. They are likely to be the product of some PCR artefact.

Nevertheless, it was re-assuring to observe that among the samples, different isolates from the same virus gave the identical "freak" sequences and that some differences were observed among the samples. Thus there was some consistency in the PCR artefact. Nonetheless, none of these sequences could be used for a phylogeny of the baculoviruses based on *polh,* due to the lack of confidence about their validity and the impossibility of ascertaining homologous sequence position with other genuine *polh* sequences.

b) Lef-8

The sequencing results from the *lef-8* fragment amplified from non-lepidopteran viruses gave more reliable results than the *polh* PCRs. The sawfly and lacewing samples, 345, 404, 413, 726 and i7, gave blast results showing that the lef-8 gene was properly targeted by the amplification. However, the extremely high score obtained for the Hemerobius NPV (table 10) showed that the sample obtained from this virus was, like the *polh* PCR, almost identical to the LdMNPV virus. This correlation leads to the

conclusion that the sample labelled 404 Hemerobius NPV was either contaminated with LdMNPV material or had been wrongly identified.

The sequences obtained from the sawfly samples were all very congruent between themselves and between rounds of PCRs from the same samples and from different samples of the same virus, for example, 345 and 413, and i7 and K14. The *lef-8* sequence from NeleNPV has high similarities with these sequences. The tblastx search of the Blast-2-sequences at NCBI gave scores of 222 & 232 and e-values of 2.E-56 & 3.E-59 for the sequences coming from the Gilpinia NPV and the Neodiprion NPVs respectively. These sequences were suitable for further phylogenetic analyses.

The sample from the cranefly *Tipula oleracea* gave a sequence that contained the primers' sequences but had otherwise no homology to the other *lef-8* sequences, even to the mosquito virus CuniNPV. This sequence could not be aligned to the lef-8 fragment used for the phylogenetic analyses and therefore was discarded. A recent blast search with this particular sequence showed that it is homologous (table 10) to a sequence from the recently completed genome of the non-occluded virus Hz-1 from *Heliothis zea* (Cheng et al., 2002). This is an extremely interesting result. It shows that this particular sample might be derived from a non-occluded "baculovirus".

c) Ac22

The *ac22* PCR amplification fragments gave extremely good sequencing results. Although fewer isolates produced positive PCRs, none of the amplifications were false-positives. Identical results were found between different rounds of PCRs and also for different samples of the same virus. The similarity search results of the sequences obtained from the sawflies baculoviruses showed that they were most similar to the *ac22* sequence of the NeseNPV sequence available in the database (table 10). This result confirmed the gene specificity of the primers and the probable authenticity of the sequences obtained from the viral isolates. These sequences were all suitable for further phylogenetic analyses.

3. Topo Walking Results

In order to acquire more genetic information about baculoviruses from non-lepidopteran hosts, a method was developed to amplify DNA fragments based on a single primer. The method was tested on a variety of samples including 47 – Actias selene NPV, 66 - Spodoptera littoralis GV, 35 – Tipula oleracea NPV, K14 - Gilpinia

hercyniae NPV, and 413 – Neodiprion sertifer NPV. Samples 47 and 66 are from lepidopteran baculoviruses and were used in these experiments as potential positive controls for the method.

Mixed results were obtained from this method. Encouraging results were obtained from the SpliGV sample. The protocol led to the amplification of a 450 bp product that could be successfully sequenced and identified as being from the right target-gene (table 11). Those results showed that the method designed to amplify a DNA fragment from one primer could work.

Table 11: Summary of the experiments done and of the results obtained for the Topo-Walking protocol.

Sample	Enzyme	Primer	PCR	Blast result: best match, score, e-value
35	BanII	polh	√ 300pb	no similarity
35	BanII	dpol	√ 400pb	Hz-1 virus NC_004156 Orf55, 44, 0.002
35	BanII	vp1054	no results	
35	BsrI	polh	no results	
35	BsrI	dpol	no results	
35	BsrI	vp1054	no results	
47	BanII	polh	no results	
47	BanII	dpol	√ 400pb	no sequence obtained
47	BanII	vp1054	no results	
47	BsrI	polh	√ 500pb	no sequence obtained
47	BsrI	dpol	√ 600pb	no sequence obtained
47	BsrI	vp1054	no results	
66	BanII	polh	√ 450pb	TnGV P06303 polh, 30, 0.013
66	BanII	dpol	no results	
66	BanII	vp1054	no results	
66	BsrI	polh	√ 450pb	no sequence obtained
66	BsrI	dpol	√ 450pb	PxGV NC-002593 ORF93 = dpol, 170, 3.e-42
66	BsrI	vp1054	no results	
K14	BsrI	dpol	√ 600pb	no sequence obtained
K14	PstI	dpol	√ 600pb	no similarity
413	BsrI	dpol	√ 450pb	no sequence obtained
413	PstI	dpol	no results	

However the results obtained from the SpliGV sample were the only ones to be entirely successful. In most cases, if a PCR fragment was obtained either from the primary PCR amplification or from the secondary nested PCR, good quality sequences proved hard to obtain. Of the few samples that produced readable sequences, only the

putative *dpol* fragment from 35-Tipula oleracea NPV cut with BanII had similarity to any sequence present in the database. This sequence is similar to Orf55 of the Hz-1 virus (table 11), like the "lef-8" fragment obtained from the same sample (table 10). Both sequences can be aligned to one another, though they were obtained using completely different methods and different primers. Since the same sequence was obtained from the same sample twice, it is probably genuine. However, it suggests that this particular sample might not be a baculovirus but some other insect virus, possibly of the same family as the Hz-1 virus. This would also explain why no other PCRs could truly work for this particular sample.

The method has proved to work for one sample. It could probably be improved by adding a cloning step after the PCR. This would certainly enhance the sequencing results. Whether this method should be preferred to a shot-gun sequencing approach remains to be determined. It was not efficient at amplifying targeted fragments from divergent viruses, which is what it was designed to achieve. The method probably needs a lot of optimisation for each individual sample. It might be efficient for researchers wanting to explore the genome of one particular virus. However, in the case of a large scale sequencing effort from numerous samples, the optimisation required for each sample might be too time consuming compared to a shot-gun sequencing approach.

4. Phylogenetic Analyses

Phylogenetic analyses were performed to elucidate the relationships of the hymenopteran baculoviruses within the *Baculoviridae*. Two genes were used in these analyses, *ac22* and *lef-8*. A complement of sequences was obtained for these 2 genes from a number of baculovirus species and isolates.

The *polh* gene was not used for the phylogenetic analyses aiming to resolve the hymenopteran baculovirus position. Two DNA sequences of this gene became available for sawflies baculovirus of two host species. However, the samples included in the phylogenetic studies did not lead to any usable sequences. Furthermore, the mosquito baculovirus CuniNPV does not have a polyhedrin gene. For the sake of data set complementarity, either CuniNPV would have had to be removed from the 3 gene data sets, or *polh* not used for the analyses. Because it is important to position the hymenopteran baculoviruses in relation to CuniNPV, this was not acceptable. It was therefore preferable not to use the *polh* gene in the analyses.

a) Alignments

Portions of DNA sequences of both the *ac22* and *lef-8* genes were aligned for further phylogenetic analyses. The taxa sampled included sequences from the 17 complete genomes available in the database (see chapter 2), from the ongoing sequencing projects of CfMNPV and NeleNPV genomes (Peter Krell and Basil Arif labs), and sequences from 8 GVs (254-Natada nararia GV, M2-Harrisina brillians GV, 384-Pieris brassicae GV, 95-Pieris rapae GV, B3-Plodia interpunctella GV, 66-Spodoptera littoralis GV, 835-Achaea janata GV), 9 lepidopteran NPVs (112-Abraxas grossulariata NPV, 789-Heliconius erato NPV, 30-Antheraea polyphemus, C2-Trichoplusia ni NPV, 344-Wiseana cervinata NPV) and 3 sawflies NPVs (345-Neodiprion sertifer NPV, 726-Neodiprion lecontei NPV, i7-Gilpinia hercyniae NPV) which I determined myself.

The complementary taxonomic sampling of the GVs and NPVs was made with 2 objectives: (i) to increase the diversity of the baculoviruses sampled and (ii) to verify the specificity of the PCR by choosing samples of viruses already present in the database such as BmNPV, OpMNPV and more recently PhopGV. Only 3 sawfly sequences were used in the analyses. Although sequences were obtained from more samples, they were very similar (345 & 413) or identical (i7, K14 & 239) to each other between samples from the same host species. Therefore to simplify the phylogenetic reconstructions by lowering the number of taxa, only one representative of each sawfly virus was used.

The DNA sequence alignments corresponded to the size of the amplified gene fragments. The *lef-8* PCR fragments produced an alignment of 513 nucleotides of which 384 were parsimony informative. For *ac22*, 250 positions of the 357 nucleotide long alignment were parsimony informative. A partition homogeneity test was performed in PAUP* to test whether these 2 genes could be combined for phylogenetic analyses. The p-value (ILD, P=0.27) showed that both data sets were congruent and could be combined into one data set.

b) Tree reconstructions

For each set of data (Lef-8, Ac22 and Lef-8+Ac22), the best-fit model of evolution was selected using MODELTEST (Posada & Crandall, 1998). All three data sets were characterised by a percentage of invariable sites (I, some positions in the alignment are constant) and a gamma parameter (G), which reflects the heterogeneity of variation rate across sites (some positions in the alignment mutate more often than

others). The nucleotide substitution model chosen for the Lef-8 data set is the HKY85 model (Hasegawa et al., 1985), which is a fairly simple 2 parameter model that allows difference in transition versus transversion mutation rates as a function of unequal base frequencies. The substitution model chosen for the Ac22 data set was the TrN model. It is a submodel of the 6 parameters general time reversible model (GTR) where only the transition rates are variable. The Lef-8+Ac22 data set also follows the TrN model. After the model had been selected, heuristic searches under maximum likelihood criteria were performed to find the best trees. These trees had different topologies from one another but basically agreed on the separation of the lepidopteran NPVs from all of the other viruses (table 12).

Table 12: Statistical differences of the ML trees obtained for the Lef-8+Ac22, Lef-8 and Ac22 data sets.

Trees¶	Lef-8+Ac22			Lef-8			Ac22		
	T1	T2	T3	TL1	TL2	TL3	TA1	TA2	TA3
Lef-8+Ac22 (TrN+G+I)†	√	√	√						
Length (-lnL)§	19507.9	19508.2	19524.6	19603.9	19588.8	19609.7	19561.0	19570.4	19543.3
Delta	BEST	0.32	16.71	96.08	80.94	101.80	53.15	62.58	35.48
KH		0.416	0.075	0.000*	0.001*	0.000*	0.005*	0.000*	0.011*
SH		0.975	0.661	0.009*	0.016*	0.004*	0.16	0.065	0.278
Lef-8 (HKY85+G+I)	√	√	√	√	√	√			
Length (-lnL)	11239.8	11240.3	11240.5	11220.5	11223.2	11238.3	11311.5	11318.7	11277.8
Delta	19.27	19.72	19.94	BEST	2.70	17.80	90.99	98.17	57.31
KH	0.142	0.13	0.118		0.394	0.071	0.000*	0.000*	0.008*
SH	0.418	0.398	0.386		0.829	0.432	0.000*	0.000*	0.021*
Ac22 (TrN +G+I)	√	√					√	√	√
Length (-lnL)	8026.6	8024.6	8043.2	8154.3	8119.6	8133.9	8001.0	8008.1	8010.4
Delta	25.54	23.54	42.15	153.32	118.53	132.85	BEST	7.09	9.40
KH	0.064	0.07	0.013*	0.000*	0.000*	0.000*		0.218	0.184
SH	0.344	0.39	0.135	0.000*	0.000*	0.000*		0.717	0.728

¶ T1, TL1, TA1: unconstrained ML trees; T2, TL2, TA2: ML trees constraint=GVs monophyletic; T3, TL3, TA3: ML trees backbone constraint= genome tree. † Best model found to fit the data set by Modeltest. § Length: ML score of the tree, Delta: length difference from the best tree, KH: Khishino-Hasegawa test under one-tailed criteria, SH: Shimodeira-Hasegawa test; KH and SH: p-value of the tree being different from the best tree; to pass: $p>0.05$. * p-value showing a significant difference to the best tree, $p<0.05$. √: trees passing both the KH and SH tests.

The topology obtained from the combined alignment (T1) showed the lepidopteran NPVs divided into 2 groups and the sawfly viruses with CuniNPV clearly separated from the GVs. However, it also showed that the GVs might be paraphyletic

(i.e. split within the tree). To test whether this hypothesis was strongly supported by the data, ML heuristic searches were performed again to find the shortest tree possible under the constraint that the GVs should be monophyletic (enforcing that the GVs should all derive from a single common ancestor). The shortest trees also showed a topology differing from that obtained with the complete genome sequences (chapter 2). Therefore phylogenies were also done by ML heuristic search enforcing the genome-tree as the backbone of the tree.

To measure the difference between the shortest topologies and those obtained with both the GV-monophyly and genome-tree constraints, KH and SH tests were performed. For each data set, the constrained trees were not significantly different from the shortest tree. However, only T1 and T2 successfully passed the tests for all three data sets (table 12). For the combined Lef-8+Ac22 data set, the tree obtained with the constraint GV-monophyletic (T2) is almost identical in length to the shortest tree (T1) (delta=0.32). This suggests that T2 represents a hypothesis for the evolution of those viruses that is equally good as that presented by T1.

In view of the current classification and phylogeny of the baculoviruses, T2 represents the most satisfactory hypothesis of the two, as the GVs are monophyletic. The shortest trees generated with either the Lef-8 or the Ac22 data sets did not show the GVs as a monophyletic group. In the case of lef-8, the GVs branch out several times from the backbone of the tree (TL1). And for ac22, the GVs include the Hymenopteran NPVs (TA1). This shows that neither data set strongly supports the GVs. However, since T2 is not significantly different from any of the shortest trees (T1, TL1, TA1; table 12), it shows that the paraphyly of the GVs is probably due to the weakness of the data.

The genome backbone constraint trees (T3, TL3, TA3) mostly changed the topology of the shortest trees (T1, TL1, TA1) in the relationships within the group II NPVs (figure 20). The trees were substantially bigger than the shortest trees but passed the KH and SH tests for their own data sets, and T3 only failed the KH test for the ac22 data set. None of those trees had the GVs as a monophyletic group.

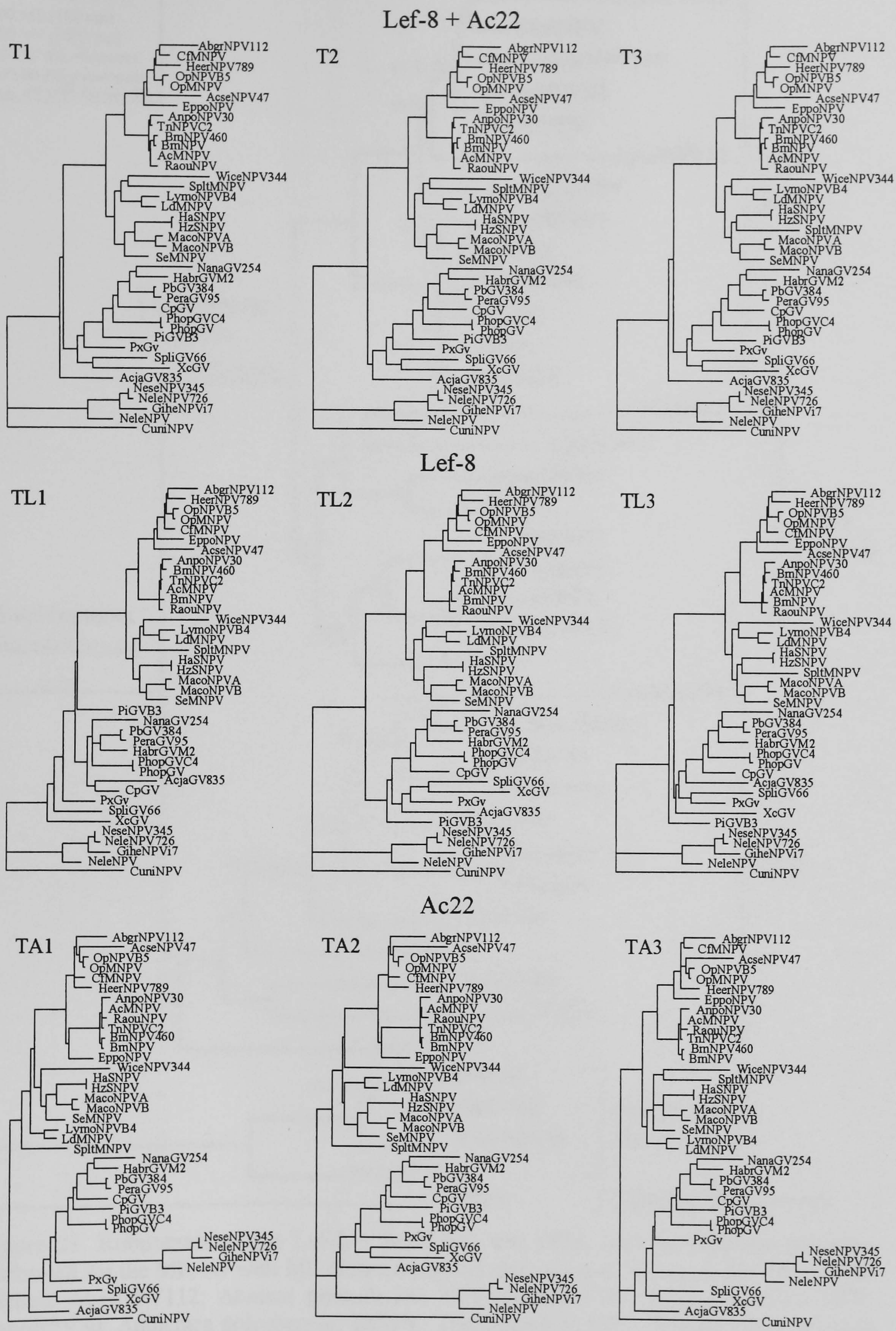


Figure 20: ML trees obtained for the Lef-8 + Ac22, Lef-8 and Ac22 data sets. T1, TL1, TA1: unconstrained trees; T2, TL2, TA2: GVs monophyletic constraint; T3, TL3, TA3: genome phylogeny backbone constraint. The trees were found by heuristic search starting with NJ and with NNI branch swapping.

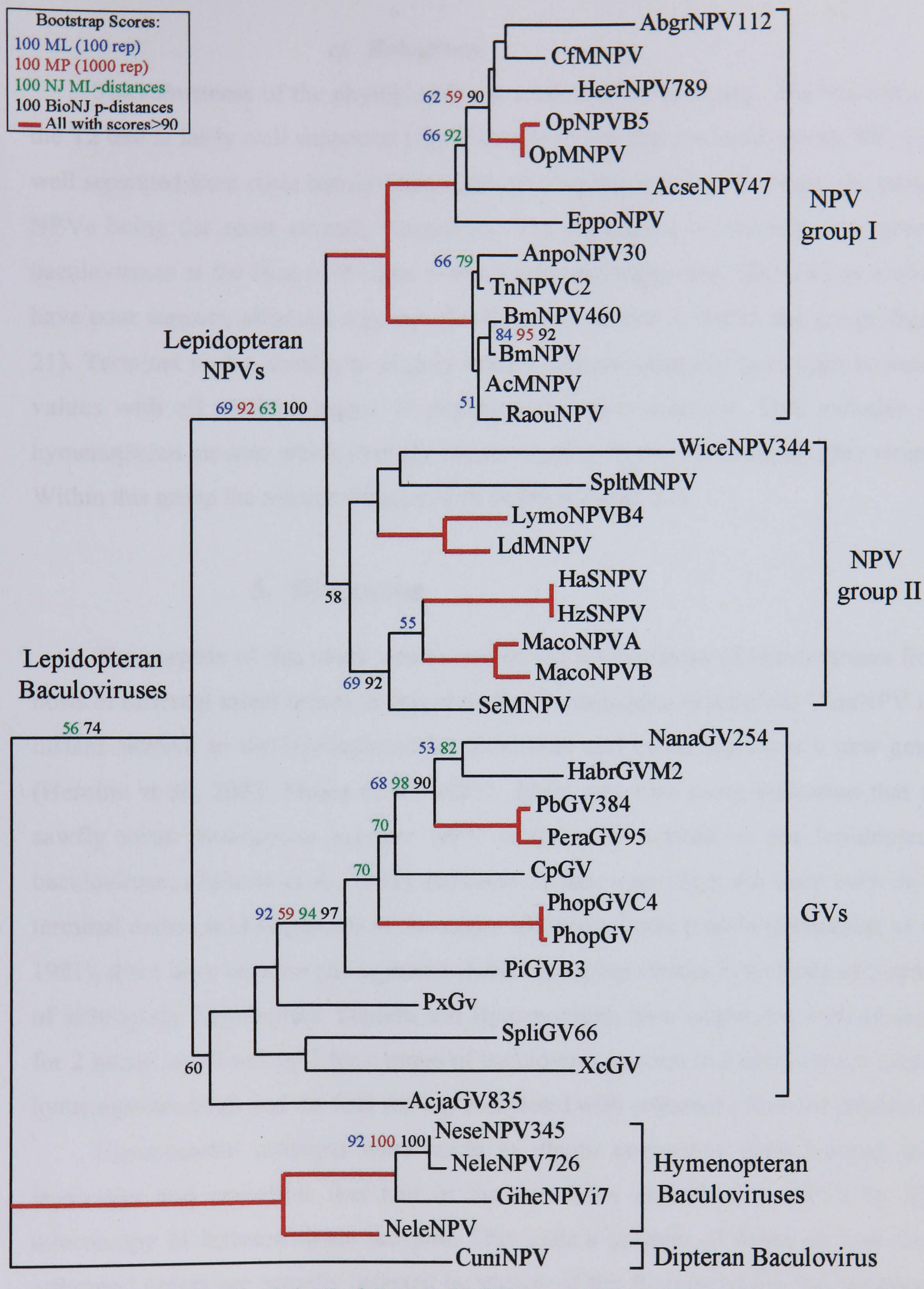


Figure 21: Robustness of the Lef-8 + Ac22 ML tree (T2). 1000 boostrap replicate were performed for the MP, NJ with ML distances, BioNJ methods and 100 repolicates for the ML method. AbgrNPV112: Abraxas grossulariata NPV; HeerNPV789: Heliconius erato NPV; AnpoNPV30: Antheraea polyphemus; TnNPV: Trichoplusia ni NPV; WiceNPV344: Wiseana cervinata NPV; NanaGV254: Natada nararia GV; HabrGVM2: Harrisina brillians GV; PbGV384: Pieris brassicae GV; PeraGV95: P.rapae GV; PiGVB3: Plodia interpunctella GV; SpliGV66: Spodoptera littoralis GV; AcjaGV835: Achaea janata GV.

c) Robustness

The robustness of the phylogenies was evaluated by bootstrap. The backbone of the T2 tree is fairly well supported (figure 21). It shows that the lepidopteran NPVs are well separated from other baculoviruses and that they are split into 2 groups, the group I NPVs being the most strongly supported. The separation of the non-lepidopteran baculoviruses at the base of the tree is also fairly well supported. The GVs as a whole have poor support, although a group of GVs seems to detach within the group (figure 21). Terminal nodes leading to closely related viruses generally have high bootstrap values with all methodologies of phylogenetic reconstruction. This includes the hymenopteran viruses, which strongly cluster together to the exclusion of other viruses. Within this group the relationships are well defined (figure 21).

5. Discussion

The purpose of this study was to outline the relationships of baculoviruses from hosts of different insect orders. It was clear that the mosquito baculovirus CuniNPV is a distant relative to the lepidopteran baculoviruses and could represent a new genus (Herniou et al., 2003, Moser et al., 2001). There was also some indication that the sawfly virus *Neodiprion sertifer* NPV is distantly related to the lepidopteran baculoviruses (Zanotto et al., 1993). However, to date apart from the early work on N-terminal amino acid sequences of the major occlusion body protein (Rohrmann et al., 1981), there have been no phylogenetic studies grouping viruses from hosts of 3 orders of arthropods: Lepidoptera, Diptera and Hymenoptera. New sequences were obtained for 2 genes: *ac22* and *lef-8* for a range of baculovirus species including viruses from 3 hymenopteran hosts and the data set was completed with sequences from the database.

Unsuccessful attempts were made to obtain sequences from isolates from lacewings and craneflies that had in the past been classified as NPVs by light microscopy of infected tissue samples. This casts a shadow of doubt on how many arthropod orders are actually infected by viruses of the *Baculoviridae*. No samples of crustacean baculoviruses were available for this study. A baculovirus of the shrimp *Paneus monodon* is still included in the ICTV tentative virus list (Blissard et al., 2000). A recent morphological description might correspond to this virus (Ramasamy et al., 2000), however, the lack of molecular sequences for this virus could still cast a shadow of doubt on its affiliation with the *Baculoviridae*, especially as another shrimp virus, the white spot syndrome virus formerly classified as a baculovirus, is in the process of

being classified to its own virus family, *Whispoviridae*, since sequences have become available (van Hulten et al., 2001, van Hulten et al., 2000).

a) Should the Baculoviridae comprise more genera?

The family *Baculoviridae* is currently split into 2 genera, *Nucleopolyhedrovirus* and *Granulovirus* (Blissard et al., 2000). So far, viruses from non-lepidopteran hosts have been classified within the NPV genus because their occlusion bodies are more similar to NPV polyhedra than to GV granules. However there has been no phylogenetic evidence that viruses from dipteran or hymenopteran hosts are part of this genus.

The phylogenetic analyses, described here, outline 4 distinct groups of baculoviruses (figure 22). The lepidopteran NPVs clearly form a discrete group to the rest of baculoviruses. The branch leading to this group is quite long (0.26) and well supported by high bootstrap values in all the analyses (figure 21). The GVs, however, appear more genetically diverse than the NPVs and the branch leading to the group is comparatively short (0.03) and poorly supported (figure 21-22). This suggests that the GVs are a much older group than the Lepidopteran NPVs, or that the sampling of the GVs was wider, or that both groups speciate at different speeds.

Furthermore, the phylogeny shows that neither the dipteran virus nor the hymenopteran viruses belong to either the GVs or the lepidopteran NPVs. The branch separating the lepidopteran from the non-lepidopteran baculoviruses is long (0.36, figure 22) and quite well supported (figure 21). The branch lengths between the mosquito virus and the sawfly viruses are also large (0.88 and 0.54). This suggests that hymenopteran and dipteran baculoviruses probably belong to distinct groups. The *Baculoviridae* appear to be clearly divided according to the classification of their hosts.

From the unrooted tree (figure 22) it is possible to envisage a scenario where the GVs could have given rise to all the NPVs (lepidopteran and non-lepidopteran). This would reconcile the phylogeny with the current classification of the baculoviruses into two genera. However, comparative genomics studies, including GVs, NPVs and CuniNPV genomes, have shown that the GVs and lepidopteran NPVs are more closely related to each other than they are to the mosquito virus (Herniou et al., 2003). Therefore this reinforces the idea that the genus *Nucleopolyhedrovirus* should probably not include viruses from a non-lepidopteran background.

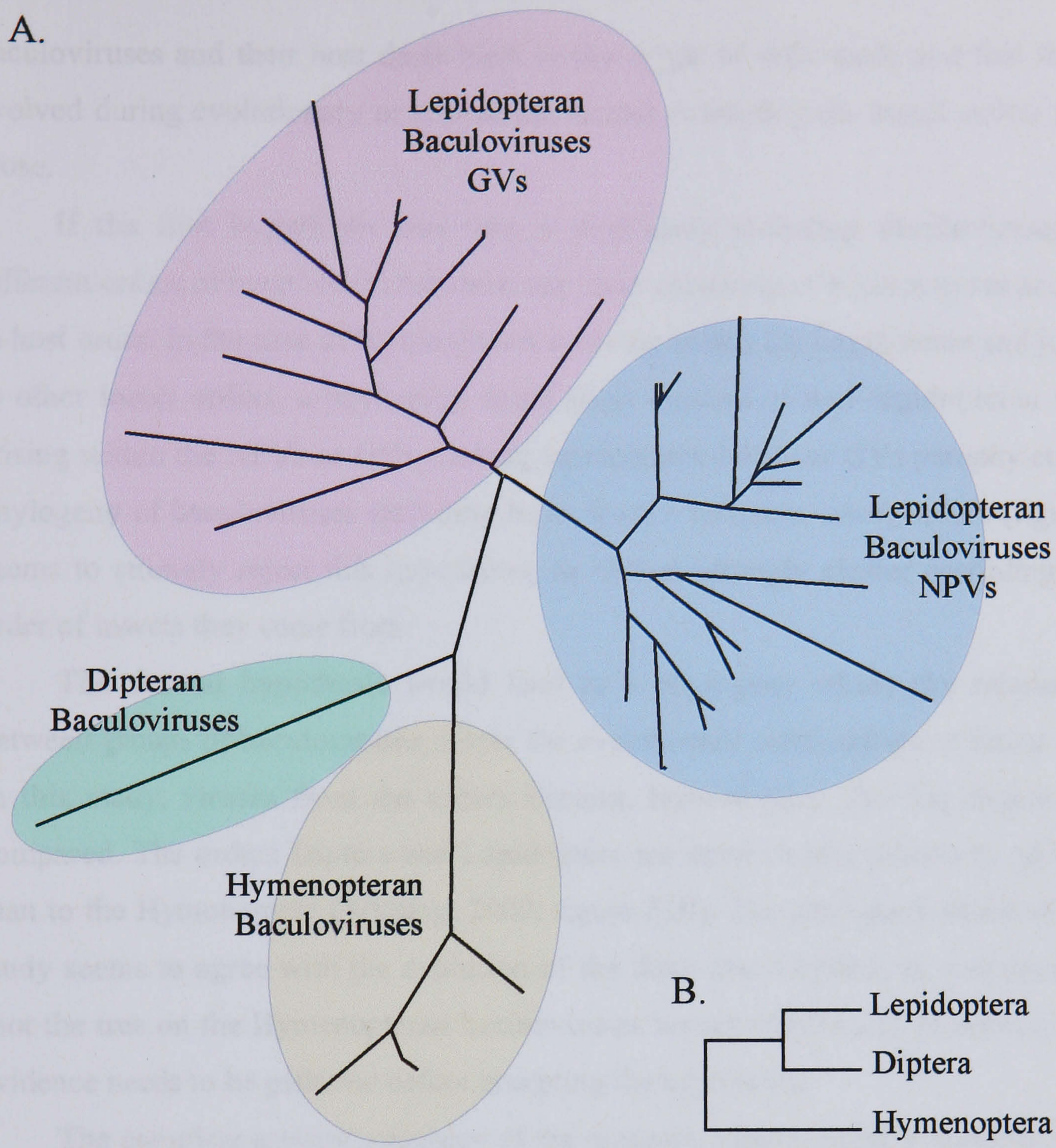


Figure 22: Evolution of the *Baculoviridae*. A. Phylogeny of the baculoviruses highlighting 4 main groups; unrooted tree (T2). B. Relationships of the 3 arthropods orders infected by the baculoviruses in A.

b) Evolution of the Baculoviridae

Baculoviruses have been described from a variety of hosts in different arthropod orders. Although some viruses, like AcMNPV, have been shown to have a wide host range within the Lepidoptera (Vail et al., 1999), most baculoviruses are thought to have a fairly narrow host range, infecting one or a few species of the same genus (Black et al., 1997, Stairs, 1971).

Two hypotheses have traditionally been proposed to explain the current host-distribution of baculoviruses (Federici, 1997). The first hypothesis states that baculoviruses could have evolved within one group of arthropods, like the Lepidoptera, and switched to other groups. The second proposes that the association between

baculoviruses and their host dates back to the origin of arthropods and that they co-evolved during evolutionary time with the viruses colonising the insect orders as they arose.

If the first hypothesis was true, a phylogeny including baculoviruses from different orders of hosts would not show any clear clustering of baculoviruses according to host order. In the case of baculoviruses evolving within the Lepidoptera and jumping to other insect orders, a phylogeny could show clusters of non–lepidopteran viruses arising within the NPVs or GVs, making lepidopteran NPVs or GVs paraphyletic. The phylogeny of baculoviruses including hosts from 3 different insect orders (figure 22) seems to strongly reject this hypothesis, as viruses strongly cluster according to the order of insects they come from.

The second hypothesis would lead to a phylogeny where the relationships between groups of baculoviruses mirror the evolutionary relationships of insect orders. In this study, viruses from the orders Diptera, Hymenoptera and Lepidoptera were compared. The orders Diptera and Lepidoptera are more closely related to each other than to the Hymenoptera (Whiting, 2002; figure 22B). The phylogeny obtained in this study seems to agree with the evolution of the three insect orders, as it is possible to root the tree on the Hymenopteran baculoviruses branch (figure 22). However, further evidence needs to be gathered before accepting the hypothesis.

The complete genome sequence of the mosquito virus CuniNPV showed that this virus does not possess a polyhedrin gene and that another protein is the major constituent of the occlusion bodies (Afonso et al., 2001, Moser et al., 2001). Polyhedrin sequences have been obtained from sawfly viruses (Rohrmann et al., 1981, Zanotto et al., 1993). This suggests that the lepidopteran and hymenopteran viruses could be more closely related to each other than to the mosquito virus. Furthermore, caterpillars and sawflies are found in the same kind of environment in the wild, terrestrial plants, and would be exposed to each others viruses whereas mosquito larvae are aquatic. This would support further the hypothesis that lepidopteran and hymenopteran baculoviruses could be more closely related. However, if the dipteran virus were more closely related to the lepidopteran viruses, it would imply that the mosquito lineage underwent a non-orthologous gene displacement for their occlusion body protein.

The pathogenesis of mosquito and sawfly viruses are more similar to each other than to that of the GVs or NPVs. These viruses only infect midgut epithelial cells (Cunningham & Entwistle, 1981, Federici, 1997, Moser et al., 2001) whereas GVs and

NPVs generally cause systemic infections. The restriction of infection to midgut epithelial cells has been proposed as an ancestral characteristic of baculoviruses (Federici, 1997). Only one lepidopteran baculovirus has been found to have a similar pathology, *Harrisina brillians GV* (HabrGV) (Federici & Stern, 1990). However, phylogenies including HabrGV showed that this virus is not basal to the GV group (figure 21) and therefore that the restriction of infection to the midgut epithelial cells is not an ancestral trait in lepidopteran baculoviruses (Bideshi et al., 2000). Therefore, it is hard to say whether the mosquito and sawfly viruses are more primitive or more derived than the lepidopteran baculoviruses based on the cell specificity of their infection.

The relationships of the deeper branches of the baculovirus phylogenies might benefit in the near future from comparative genomics analyses. If sawfly virus genomes were found to share more genes with the lepidopteran baculoviruses, then CuniNPV would remain at the base of the tree. If they share more genes with CuniNPV, then the hymenopteran and dipteran viruses could be grouped together to the exclusion of the lepidopteran baculoviruses. However, if CuniNPV shares more genes with the lepidopteran baculoviruses, then the hymenopteran baculoviruses could be the more ancestral lineage. This last option would favour the theory of co-evolution between the *Baculoviridae* and the Arthropoda.

c) Taxonomic propositions

This study shows that under the current classification, CuniNPV and the sawfly NPVs should be removed from the genus *Nucleopolyhedrovirus* and classified under the section "unclassified baculoviruses". It would also support a taxonomic proposition to create one or two new genera of baculoviruses. The number of new genera would depend on further evidence to cluster together or keep apart the dipteran and hymenopteran baculoviruses.

The study supports the status of distinct species for *Gilpinia hercyniae NPV*. BLAST search results showed that NeseNPV-345 &-413 as well as NeleNPV-726 are isolates of the species *Neodiprion sertifer NPV* (taxonomic code: 00.006.0.01.017.). The other virus isolated from *N. lecontei* should still be classified as a separate species *Neodiprion lecontei NPV* (taxonomic code: 00.006.0.01.323.).

D. Phylogeny of Lepidopteran baculoviruses

In this section, the focus is only on baculoviruses from lepidopteran hosts. The previous section has shown that the baculoviruses from non-lepidopteran hosts, available to this study, are distantly related to both the lepidopteran NPVs and the GVs. The analyses so far have shown that both groups are well defined and non-overlapping (monophyletic). The NPVs and the GVs from lepidopteran hosts can be used as each other's outgroup. It was therefore possible to remove the non-lepidopteran taxa to simplify the analyses.

The purpose of this section was to explore the diversity of lepidopteran baculoviruses by sampling as many taxa as possible, from as many host species as possible, and making a particular effort to sample from a wide range of lepidopteran families. The aims are (1) to reconstruct gene genealogies of lepidopteran baculoviruses (2) to assess the congruence between different gene trees to make hypotheses of baculovirus evolution (3) to explore the relationships between viral isolates in order to classify them into strains and virus species.

1. Taxa and Sequence Sampling

A total of 141 virus samples from 97 different species of Lepidoptera were available for this study (appendix 2). This included baculoviruses from 19 lepidopteran families spanning 12 superfamilies. For all the samples, DNA extraction was performed, followed by PCR amplification and direct sequencing of positives. The genes targeted for this study of the diversity of lepidopteran baculoviruses included *polyhedrin* (primers PF+PR), as this gene is the best represented baculovirus gene in the databases. *Lef-8* was chosen as the second gene to use. This time, the PCR fragment was amplified with the L8F1 and L8R1 primers, this is a larger and non-overlapping fragment to the one used in section C.

Sequences for one or both genes were obtained from 98 of the 141 samples, a sequencing success rate of 69.5%. They represented viruses from 83 different host species spanning all 19 lepidopteran families (table 13). This sequencing rate was quite encouraging overall, considering the range of sample quality. There are a number of reasons why a sample would not work: (i) the original sample could have been very dilute; (ii) the DNA could have been degraded; (iii) the sample could have been

wrongly labelled as an NPV. A few of the samples tested were labelled NPV+CPV, for a possible baculovirus and cypovirus co-infection. Some of these samples gave positive results for baculovirus, but others proved negative. Those latter samples could contain cypoviruses only. All the 24 GV samples examined gave positive sequencing results (table 13), showing that although they are much smaller and harder to see under the light microscope than the NPVs they might be easier to identify as *Granulovirus*.

Table 13: List of the lepidopteran baculoviruses used in the *polyhedrin* and *lef-8* phylogenetic analyses

Virus Name †	Host Superfamily	Host Family	Virus type	Acc#	*polh*	*lef-8*	Collector	Country
Bombyx mori NPV	Bombycoidea	Bombycidae	NPV	L33180	1	1		
Bombyx mori NPV460	Bombycoidea	Bombycidae	NPV	460	1	1	P. Seraphim	Poland
Lasiocampa quercus NPV202	Bombycoidea	Lasiocampidae	NPV	202	1	1	IVEM	UK
Malacosoma californicum pluviale NPVk3	Bombycoidea	Lasiocampidae	NPV	k3	1	1	J. Myers	Canada
Malacosoma Disstria NPV	Bombycoidea	Lasiocampidae	NPV	U61732	1			
Malacosoma neustria NPV	Bombycoidea	Lasiocampidae	NPV	X55658	1			
Malacosoma neustria NPVk12	Bombycoidea	Lasiocampidae	NPV	k12	1	1	J. Myers	Canada
Pachypasa papyri NPV436	Bombycoidea	Lasiocampidae	NPV	436	1	1	Mrs Aloo	Kenya
Trabala vishnou NPV195	Bombycoidea	Lasiocampidae	NPV	195	1	1	IVEM	UK
Actias selene NPV47	Bombycoidea	Saturniidae	NPV	47	1	1	IVEM	UK
Antheraea pernyi NPV	Bombycoidea	Saturniidae	NPV	M57429	1	1		
Antheraea polyphemus NPV30	Bombycoidea	Saturniidae	NPV	30	1		IVEM	UK
Attacus ricini NPV	Bombycoidea	Saturniidae	NPV	S68462	1			
Lomonia obliqua NPV	Bombycoidea	Saturniidae	NPV	AF23690	1		Wolff	Brazil
Samia ricini NPV28	Bombycoidea	Saturniidae	NPV	28	1	1	IVEM	UK
Phthorimaea operculella GV	Gelehioidea	Gelechiidae	GV	AF499596	1	1	L. Croizier	Morocco
Phthorimaea operculella GVc3	Gelehioidea	Gelechiidae	GV	c3	1	1	J. Cory	Nicaragua
Phthorimaea operculella GVc4	Gelehioidea	Gelechiidae	GV	c4	1	1	J. Cory	Nicaragua
Abraxas grossulariata NPV112	Geometroidea	Geometridae	NPV	112	1	1	IVEM	UK
Buzura edwardsii NPV505	Geometroidea	Geometridae	NPV	505	1	1	A. Weaving	Rodhesia
Buzura supressaria NPV	Geometroidea	Geometridae	NPV	X70844	1		Z. Hu	China
Ectropis obliqua NPV	Geometroidea	Geometridae	NPV	U95104	1			
Erannis defolaria NPV174	Geometroidea	Geometridae	NPV	174	1	1	IVEM	UK
Idaea virgularia NPV402	Geometroidea	Geometridae	NPV	402	1	1	IVEM	UK
Operopthera brumata NPV291	Geometroidea	Geometridae	NPV	291	1	1	IVEM	UK
Operopthera brumata NPVi4	Geometroidea	Geometridae	NPV	i 4	1		IVEM	UK
Operopthera brumata NPVi5	Geometroidea	Geometridae	NPV	i 5	1		IVEM	UK
Wiseana cervinata GV342	Hepialoidea	Hepialidae	GV	342	1	1	IVEM	New Zeland
Wiseana cervinata NPV344	Hepialoidea	Hepialidae	NPV	344	1	1	IVEM	New Zeland
Wiseana signata NPV	Hepialoidea	Hepialidae	NPV	AF016916	1			New Zeland
Ammalo insulata NPV175	Noctuoidea	Arctiidae	NPV	175	1	1	IVEM	UK
Amsacta albistriga NPV	Noctuoidea	Arctiidae	NPV	AF118850	1			
Diacrisia obliqua NPV798	Noctuoidea	Arctiidae	NPV	798	1	1	Dr Pawar	India
Hyphantria cunea GV228	Noctuoidea	Arctiidae	GV	228	1	1	IVEM	UK

Hyphantria cunea NPV	Noctuoidea	Arctiidae	NPV	D14573	1			
Hyphantria cunea NPVn3	Noctuoidea	Arctiidae	NPV	N3	1			Japan
Spilosoma imparilis NPVn2	Noctuoidea	Arctiidae	NPV	N2	1	1		Japan
Spilosoma lutea GV126	Noctuoidea	Arctiidae	GV	126	1	1	IVEM	UK
Utetheisa pulchella NPV142	Noctuoidea	Arctiidae	NPV	142	1		IVEM	UK
Aroa discalis NPV63	Noctuoidea	Lymantriidae	NPV	63	1	1	K. Smith	Nigeria
Euproctis chrysorrhoea NPVa10	Noctuoidea	Lymantriidae	NPV	a10	1	1	IVEM	UK
Euproctis similis NPV768	Noctuoidea	Lymantriidae	NPV	768	1	1	H. Evans	UK
Lymantria dispar NPV	Noctuoidea	Lymantriidae	NPV	AF081810	1	1		
Lymantria monacha NPVb4	Noctuoidea	Lymantriidae	NPV	b4	1	1	IVEM	UK
Lymantria nanayi NPV750	Noctuoidea	Lymantriidae	NPV	750	1	1	H. Roberts	N.Guinea
Lymantria obfuscata NPV193	Noctuoidea	Lymantriidae	NPV	193	1	1	IVEM	Kashmir
Orgyia anartoides NPV	Noctuoidea	Lymantriidae	NPV	AF068188	1		R. Teakle	Australia
Orgyia antiqua NPVa9	Noctuoidea	Lymantriidae	NPV	a9	1	1	IVEM	UK
Orgyia leusostigma NPVb6	Noctuoidea	Lymantriidae	NPV	b6	1	1	IVEM	UK
Orgyia mixta NPV67	Noctuoidea	Lymantriidae	NPV	67	1	1	K. Smith	Nigeria
Orgyia pseudotsugata MNPV	Noctuoidea	Lymantriidae	NPV	U75930	1	1		
Orgyia pseudotsugata NPVb5	Noctuoidea	Lymantriidae	NPV	b5	1		IVEM	UK
Orgyia pseudotsugata SNPV	Noctuoidea	Lymantriidae	NPV	M32433	1			
Perina nuda NPV	Noctuoidea	Lymantriidae	NPV	U22824	1			
Achaea faber NPV64	Noctuoidea	Noctuidae	NPV	64	1	1	K. Smith	Nigeria
Achaea janata GV835	Noctuoidea	Noctuidae	GV	835	1		Dr Pawar	India
Agrotis segetum NPVi6	Noctuoidea	Noctuidae	NPV	i 6	1		IVEM	UK
Agrotis segetum NPVk4	Noctuoidea	Noctuidae	NPV	k4	1		IVEM	UK
Anagrapha flacifera NPVk2	Noctuoidea	Noctuidae	NPV	k2	1	1	Hoffman	USA
Anomis sabulifera NPV14	Noctuoidea	Noctuidae	NPV	14	1	1	IVEM	UK
Anticarsia gemmatalis NPV	Noctuoidea	Noctuidae	NPV	Y17753	1			
Anticarsia gemmatalis NPVb10	Noctuoidea	Noctuidae	NPV	b10		1	IVEM	UK
Autographa californica NPV	Noctuoidea	Noctuidae	NPV	L22858	1	1		
Autographa californica NPVk1	Noctuoidea	Noctuidae	NPV	k1	1	1	Hoffman	USA
Cerapteryx graminis NPV a5	Noctuoidea	Noctuidae	NPV	a5	1	1	IVEM	UK
Cirphis unipuncta NPV172	Noctuoidea	Noctuidae	NPV	172	1		IVEM	UK
Euplexia lucipara GV248	Noctuoidea	Noctuidae	GV	248	1	1	IVEM	UK
Helicoverpa armigera SNPV	Noctuoidea	Noctuidae	NPV	AF271059	1	1		
Helicoverpa zea SNPV	Noctuoidea	Noctuidae	NPV	AF334030	1	1		
Lacanobia oleracea GVa6	Noctuoidea	Noctuidae	GV	a6	1	1	I. Smith	UK
Mamestra brassicae NPV	Noctuoidea	Noctuidae	NPV	M20927	1		M. Summers	USA
Mamestra brassicae NPV274	Noctuoidea	Noctuidae	NPV	274	1	1	IVEM	UK
Mamestra brassicae NPVk9	Noctuoidea	Noctuidae	NPV	k9	1	1	IVEM	UK
Mamestra configurata NPVA	Noctuoidea	Noctuidae	NPV	AF467808	1	1		Canada
Mamestra configurata NPVB	Noctuoidea	Noctuidae	NPV	AY126275	1	1		Canada
Melanchra persicariae GV36	Noctuoidea	Noctuidae	GV	26	1	1	K. Smith	UK
Mythimna separata NPV	Noctuoidea	Noctuidae	NPV	U30302	1			
Mythimna separata NPVa2	Noctuoidea	Noctuidae	NPV	a2	1	1	J. Cory	Japan
Mythimna separata NPVa3	Noctuoidea	Noctuidae	NPV	a3	1	1	J. Cory	China
Mythimna separata NPVa4	Noctuoidea	Noctuidae	NPV	a4	1	1	J. Cory	India
Mythimna unipuntata NPVa1	Noctuoidea	Noctuidae	NPV	a1	1	1	J. Cory	America
Panolis flammea NPV	Noctuoidea	Noctuidae	NPV	D00437	1			
Panolis flammea NPVi3	Noctuoidea	Noctuidae	NPV	i 3	1	1	IVEM	UK
Plusia gamma NPV287	Noctuoidea	Noctuidae	NPV	287	1	1	IVEM	UK
Plusia orichalcea NPV	Noctuoidea	Noctuidae	NPV	AF01988	1			

Rachiplusia ou NPV	Noctuoidea	Noctuidae	NPV	AY145471	1	1		
Spodoptera exempta NPVk11	Noctuoidea	Noctuidae	NPV	k11	1	1	IVEM	UK
Spodoptera exigua MNPV	Noctuoidea	Noctuidae	NPV	AF169823	1	1		
Spodoptera frugiperda NPV	Noctuoidea	Noctuidae	NPV	J04333	1			
Spodoptera frugiperda NPVk10	Noctuoidea	Noctuidae	NPV	k10	1	1	IVEM	UK
Spodoptera littoralis GV66	Noctuoidea	Noctuidae	GV	66	1	1		Ceylon
Spodoptera littoralis NPV	Noctuoidea	Noctuidae	NPV	D01017	1		x	x
Spodoptera littoralis NPV440	Noctuoidea	Noctuidae	NPV	440	1	1	IVEM	UK
Spodoptera litura GV76	Noctuoidea	Noctuidae	GV	766	1	1	T. Sankaran	SriLanka
Spodoptera litura NPV	Noctuoidea	Noctuidae	NPV	X94437	1			
Spodoptera litura NPVg	Noctuoidea	Noctuidae	NPV	AF325155	1	1		
Spodoptera sunia NPVk7	Noctuoidea	Noctuidae	NPV	k7	1		J. Cory	Nicaragua
Spodoptera sunia NPVk8	Noctuoidea	Noctuidae	NPV	k8	1	1	J. Cory	Nicaragua
Thysanoplusia orichalcea NPV1	Noctuoidea	Noctuidae	NPV	AF169480	1			
Thysanoplusia orichalcea NPV2	Noctuoidea	Noctuidae	NPV		1		Cheng, 1998	
Trichoplusia ni GV	Noctuoidea	Noctuidae	GV	K02910	1			
Trichoplusia ni NPVc2	Noctuoidea	Noctuidae	NPV	c2	1	1	IVEM	UK
Trichoplusia ni SNPV	Noctuoidea	Noctuidae	NPV	AF093405	1			
Trichoplusia orichalcea NPVb9	Noctuoidea	Noctuidae	NPV	b9	1		IVEM	UK
Xestia c-nigrum GV	Noctuoidea	Noctuidae	GV	AF162221	1	1		
Cynosarga chrysolopa NPV302	Noctuoidea	Notodontidae	NPV	302	1	1	IVEM	UK
Aglais urticae NPVa7	Papilionidea	Nymphalidae	NPV	a7	1	1	IVEM	UK
Agraulis vanillae NPV779	Papilionidea	Nymphalidae	NPV	779	1		IVEM	UK
Helioconius erato NPV789	Papilionidea	Nymphalidae	NPV	789	1	1	IVEM	UK
Junonia coenia GV19	Papilionidea	Nymphalidae	GV	19	1	1	IVEM	UK
Nymphalis io NPV353	Papilionidea	Nymphalidae	NPV	353	1	1	IVEM	UK
Polygonia c-album NPV380	Papilionidea	Nymphalidae	NPV	380	1	1	IVEM	UK
Vanessa atalanta NPV54	Papilionidea	Nymphalidae	NPV	54	1	1	IVEM	UK
Vanessa cardui NPV367	Papilionidea	Nymphalidae	NPV	367	1	1	IVEM	UK
Aporia crataegi NPV106	Papilionidea	Pieridae	NPV	106	1	1	IVEM	UK
Artogeia napi GV242	Papilionidea	Pieridae	GV	242	1	1	IVEM	UK
Catopsila pomona NPV78	Papilionidea	Pieridae	NPV	78	1	1	K. Smith	N.Guinea
Pieris brassicae GV	Papilionidea	Pieridae	GV	X02498	1			
Pieris brassicae GV384	Papilionidea	Pieridae	GV	384	1	1	K. Smith	UK
Pieris rapae GV95	Papilionidea	Pieridae	GV	95	1	1	K. Smith	UK
Diatrea saccharalis GVk6	Pyraloidea	Pyralidae	GV	k6	1		IVEM	UK
Ephestia cautella NPVc5	Pyraloidea	Pyralidae	NPV	c5	1	1	IVEM	UK
Plodia interpunctella GVb3	Pyraloidea	Pyralidae	GV	b3	1	1	IVEM	UK
Smerinthus ocellata NPV819	Sphingoidea	Sphingidae	NPV	819	1		IVEM	UK
Sphinx ligustri NPV334	Sphingoidea	Sphingidae	NPV	334	1		IVEM	UK
Sphinx ligustri NPV362	Sphingoidea	Sphingidae	NPV	362	1	1	IVEM	UK
Kotochalia junodi NPV205	Tineoidea	Psychidae	NPV	205	1	1	IVEM	UK
Kotochalia junodi NPV32	Tineoidea	Psychidae	NPV	32	1	1	IVEM	UK
Mahasena corbetti NPV595	Tineoidea	Psychidae	NPV	595	1		IVEM	UK
Mahasena corbetti NPV692	Tineoidea	Psychidae	NPV	692	1	1	S. Shah	Malaysia
Mahasena corbetti NPV699	Tineoidea	Psychidae	NPV	699	1		S. Shah	Malaysia
Adoxophyes orana GV	Tortricoidea	Tortricidae	GV	AF337646	1			
Archips cerasivora NPV	Tortricoidea	Tortricidae	NPV	U40834	1			
Cacoecia murinana GV37	Tortricoidea	Tortricidae	GV	37	1	1	J. Weiser	
Choristoneura furamifera DEFNPV	Tortricoidea	Tortricidae	NPV	U78194	1		B. Arif	Canada
Choristoneura furamifera GV	Tortricoidea	Tortricidae	GV	AF058690	1			

Choristoneura furamifera MNPV	Tortricoidea	Tortricidae	NPV		1	1	B. Arif	Canada
Choristoneura rosaceana NPV	Tortricoidea	Tortricidae	NPV	U91940	1			
Cryptophlebia leucotreta GV	Tortricoidea	Tortricidae	GV	X79569	1		J. Jehle	
Cydia pomonella GV	Tortricoidea	Tortricidae	GV	U53466	1	1	N. Crook	Mexico
Epyphias posvittana MNPV	Tortricoidea	Tortricidae	NPV	AY043265	1	1		
Homona coffearia GV744	Tortricoidea	Tortricidae	GV	744	1	1	D. Shaw	N.Guinea
Homona coffearia GV745	Tortricoidea	Tortricidae	GV	745	1		D. Shaw	N.Guinea
Anthophila fabricana GV790	Yponomeutoidea	Glyphipterigidae	GV	790	1		IVEM	UK
Plutella xylostella GV	Yponomeutoidea	Plutellidae	GV	AF270937	1	1		
Darna trima GV545	Zygaenoidea	Limacodidae	GV	545	1	1	Dr Munroe	Malaysia
Natada nararia GV254	Zygaenoidea	Limacodidae	GV	254	1	1	K. Smith	UK
Harrisina brillans GV	Zygaenoidea	Zygaenidae	GV	AF142425	1		B. Federici	USA
Harrisina brillans GVm2	Zygaenoidea	Zygaenidae	GV	M2	1	1	B. Federici	USA

† Virus names correspond in regular font to the sequences obtained in this study, in italics to sequences from geneBank or publications, in bold to sequences from genome projects.

2. Phylogenetic Analyses of the Polyhedrin Gene Fragment

a) Sequences and alignment

The PCR amplification and direct sequencing worked extremely well for the *polh* gene fragment. In total, sequences were obtained from 97 samples including 24 GVs and 77 NPVs (table 13). The data set was completed with 51 DNA sequences from the database (table 13). The 148 sequences were aligned in MacClade (Maddison & Maddison, 2000). This alignment was easy to do, as the *polh* gene is well conserved among lepidopteran baculoviruses, with few triplet insertions or deletions (appendix 3). Of the alignment 537 nucleotide-positions, 25.5% (137) were constant, the same nucleotide being conserved across all taxa. Of the 400 remaining variable positions, 359 (66.8%) were parsimony informative.

b) Model of Sequence Evolution

The best-fit model of evolution for the 148 *polh* sequences was selected using MODELTEST (Posada & Crandall, 1998). A GTR+G+I model was selected for the data set. Specific values were given for the percentage of invariable sites (I) and the gamma parameter (G). The nucleotide substitution model chosen for the *polh* data set is a general time reversible (GTR) model (Rodriguez et al., 1990), where all 6 parameters of transition or transversion mutation rates have different values (A<>T ≠ A<>C ≠ A<>G ≠ T<>C ≠ T<>G ≠ C<>G).

c) Polh Phylogeny

The *polh* phylogeny was obtained under maximum likelihood criteria by heuristic search with NNI branch swapping. The tree (figure 23) showed a clear division between the NPVs and the GVs, which was supported by high bootstrap values for the different methodologies tested. Terminal branches also obtained high bootstrap scores, regrouping strains and closely related viruses. However the tree showed a clear lack of resolution at the intermediate level. Its structure was not supported by bootstrap, apart from the group I NPVs, which obtained fairly high bootstrap scores for all the methods. The group II NPV has completely collapsed, showing that there is little structure in the basal branches of the lepidopteran NPVs for this gene phylogeny.

A striking feature of this *polh* phylogeny is the position of AcMNPV. This virus, the archetype of the group I NPVs (see Chapter 2), has completely fallen out of the group I clade and is confined deep in the phylogeny with group II NPVs viruses. To test the robustness of this topology (figure 23), searches were performed with a number of topological constraints and the trees obtained were compared to the best tree using KH and SH tests. The constraints tested were group II NPVs including: AcMNPV being monophyletic as opposed to paraphyletic, AcMNPV within the group I NPVs, AcMNPV with BmNPV and RaouNPV, or the genome tree backbone (AcMNPV also sister group to BmNPV). Here AcMNPV was taken as a clade including the genome sequence and affiliated strains. The trees were obtained in a similar manner as before, only imposing the topological constraints. The significance of the topological differences were measured by SH and KH tests (table 14).

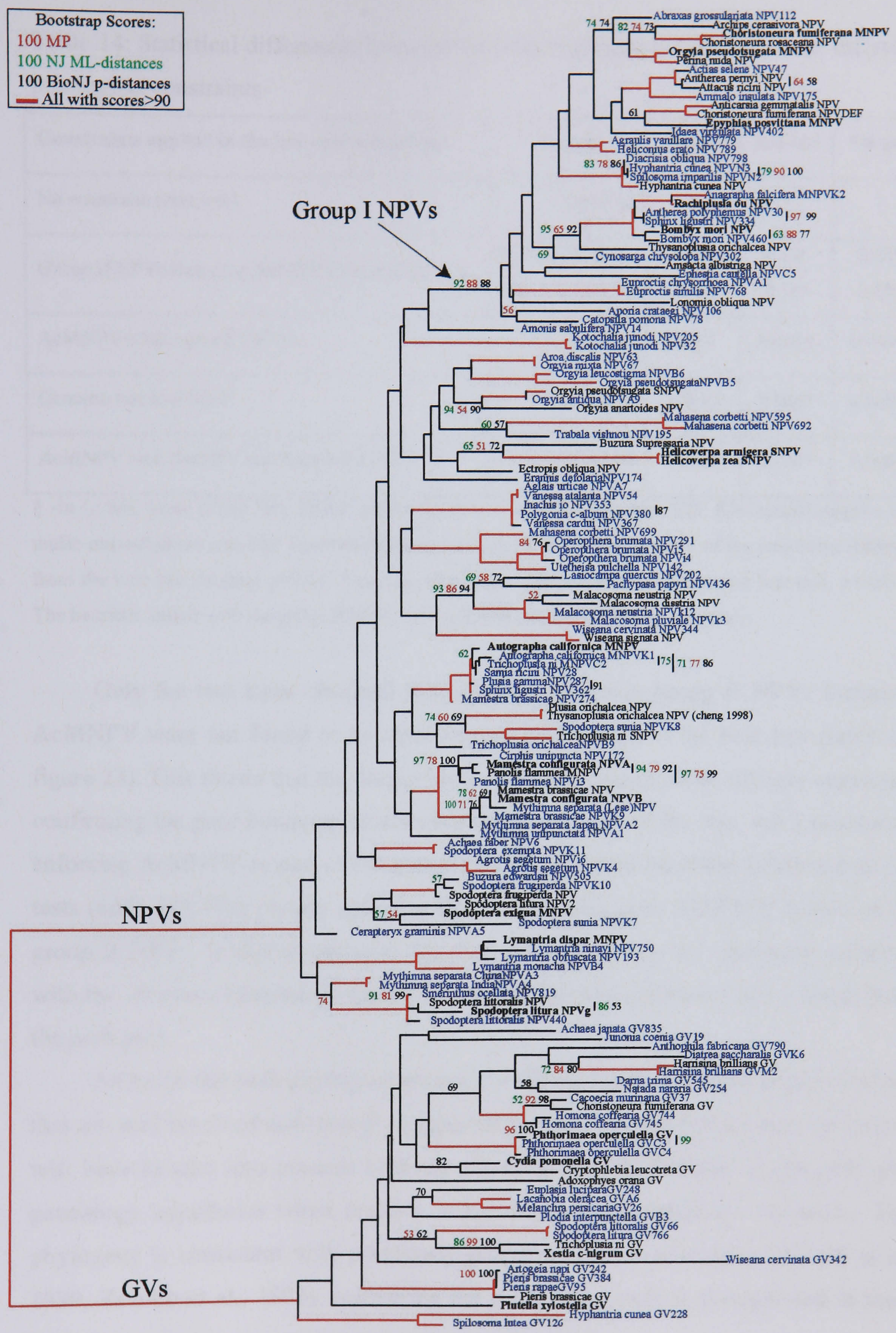


Figure 23: *Polyhedrin* Phylogeny. ML tree, 1000 boostrap replicates were performed for the MP, NJ with ML distances and BioNJ methods. **Orgyia pseudotsugata MNPV:** in bold sequences from complete genomes, Perina nuda NPV: in black sequences from GeneBank, Actias selene NPV47: in blue sequences acquired in this study.

Table 14: Statistical differences between the *polh* phylogenies obtained under different topological constraints.

Constraints applied in the tree reconstructions	**-ln L §**	**Delta**	**KH-test**	**SH-test**
No constraint (best tree)	23019.82	0	1	1
Group II NPVs including AcMNPV monophyletic †	tree 1: 23048.88 tree 2: 23048.92	29.06 29.10	0.186 0.183	0.485 0.486
AcMNPV within group I NPVs	23113.90	94.08	0.000*	0.029*
Genome tree backbone	23213.74	193.92	0.000*	0.000*
AcMNPV with BmNPV and RaouNPV clade	23185.16	165.34	0.000*	0.000*

§ -ln L: ML score of the tree, Delta: length difference from the best tree, KH: Khishino-Hasegawa test under one-tailed criteria, SH: Shimodeira-Hasegawa test; KH and SH: p-value of the tree being different from the best tree; to pass: $p>0.05$. * p-value showing a significant difference to the best tree, $p<0.05$. † The heuristic search with the group II NPVs monophyletic constraint yielded two trees.

Only the two trees obtained with the monophyletic group II NPVs including AcMNPV were not found to be significantly different from the best tree (table 14, figure 23). This shows that the deeper level of the phylogeny is not strongly supported, confirming the poor bootstrap scores obtained in this part of the tree. All 3 constraints enforcing AcMNPV as part of the group I NPVs produced trees that failed to pass the tests (table 14). This clearly indicates that for the *polh* gene AcMNPV is part of the group II NPVs. It also emphasizes the discrepancy between the phylogeny obtained with the 30 genes common to the baculoviruses and that obtained from a single gene, the *polh* gene.

Although the *polh* phylogeny shows a weak backbone there are parts of the tree that are well resolved and strongly supported. Further interpretation based on the tree will have to take into account both the strengths and weaknesses of the *polh* gene genealogy hypothesis when making inferences on virus species / isolates. This phylogeny is consistent with previously published *polh* phylogenies (Bulach et al., 1999, Zanotto et al., 1993), recovering the groups that were well-supported in those trees.

3. Phylogenetic Analyses of the Lef-8 Gene Fragment

a) Sequences and alignment

The PCR amplification and direct sequencing did not work as well for the *lef-8* gene fragment as for the *polh* fragment. Of the 97 samples for which a *polh* sequence was obtained, only 79 gave a *lef-8* PCR amplification. Only one of the samples, Anticarsia gemmatalis NPV-b10, successfully amplified *lef-8*, but not *polh*. The PCR amplification products showed quite a large variation in size, with fragments of 750 to 900 bp. The sequencing of these *lef-8* fragments proved in some cases more complicated than expected. For a majority of samples sequencing the fragments was relatively straightforward. But for other samples, puzzling results were obtained time and time again. For these latter fragments, the forward and reverse sequencing primers produced sequences that were the same. This should not happen, but these results were repeated several times and with sufficient controls to guarantee that they did not result from errors in the sequencing reactions set up. Some secondary structures within these fragments were thought to have caused these sequencing artefacts but no further evidence within these sequences was found. The consequences were that 20 taxa had shorter sequences, missing the 3' and/or 5' ends of the target *lef-8* fragment.

In total, sequences were obtained from 80 samples, including 20 GVs and 60 NPVs (table 13). The data set was completed with 17 baculovirus *lef-8* DNA sequences from the database (table 13). The 97 sequences were aligned in MacClade (Maddison & Maddison, 2000). The alignment was not as straightforward as that of the *polh* fragment. Reflecting the size variation of the PCR products, large insertions had to be introduced to allow for the alignment of more conserved regions (appendix 3). The largest insertion was 279 nucleotides long (equivalent to 93 amino acids) for Euplasia lucipara GV-248. Because the sequences were less conserved and therefore difficult to align around the insertion/deletion points, the relationships of homology of nucleotides at these positions were not convincing. To remove the noise that such unaligned positions could bring to the phylogenetic reconstruction, the alignment was trimmed to keep only the most conserved region of this target *lef-8* gene fragment. The alignment was therefore reduced by 36% from 1017 to 654 nucleotides (appendix 3). With this alignment, only 13.8% (90) of nucleotide positions were constant, which was 11.7% less conserved than the *polh* fragment. However, of the 564 variable positions, 514 (78.6%) were parsimony informative, which is 11.8% more informative than *polh*.

Although the *lef-8* sequences were more difficult to acquire and align than the *polh* sequences, it appeared that they might be phylogenetically more informative.

b) Model of Sequence Evolution and Lef-8 Phylogeny

As for the *polh* alignment, MODELTEST (Posada & Crandall, 1998) selected a GTR+G+I model for the 98 sequences of the *lef-8* data set. The *lef-8* phylogeny was obtained under maximum likelihood criteria by heuristic search with NNI branch swapping. The tree (figure 24) showed some lack of resolution at the intermediate level. It obtained fairly good bootstrap scores for the division between the NPVs and the GVs as well as for the group I NPVs. Terminal branches also obtained high bootstrap scores, regrouping strains and closely related viruses. The group II NPVs was paraphyletic, appearing to be split into 2 main groups.

To further test the robustness of this topology, the search was repeated with two constraints. The first was the monophyly of the group II NPVs. The tree obtained with this constraint passed both the KH ($p=0.227$) and SH ($p=0.444$) tests. The difference in length to the best tree was 9.72. This result emphasized the weakness of the phylogeny at intermediary levels, as shown by the low bootstrap scores obtained for those branches. The second constraint tested was the genome tree backbone. The tree obtained with this imposed backbone topology failed to pass both the SH and KH test ($p<0.000$). The tree length difference compared to the best tree was 83.73. This negative result highlights again the differences between the complete genome phylogeny and single gene fragment phylogeny.

Although the *lef-8* phylogeny did not present a much stronger backbone than that of the *polh* phylogeny, there are strongly supported parts of the tree on which hypotheses of virus relationships can be inferred. The following sections will look at the congruence at the tips of the trees of both gene phylogenies to infer taxonomic proposals and will look at the overall congruence or incongruence of the tree to try to make inferences about lepidopteran baculovirus evolution.

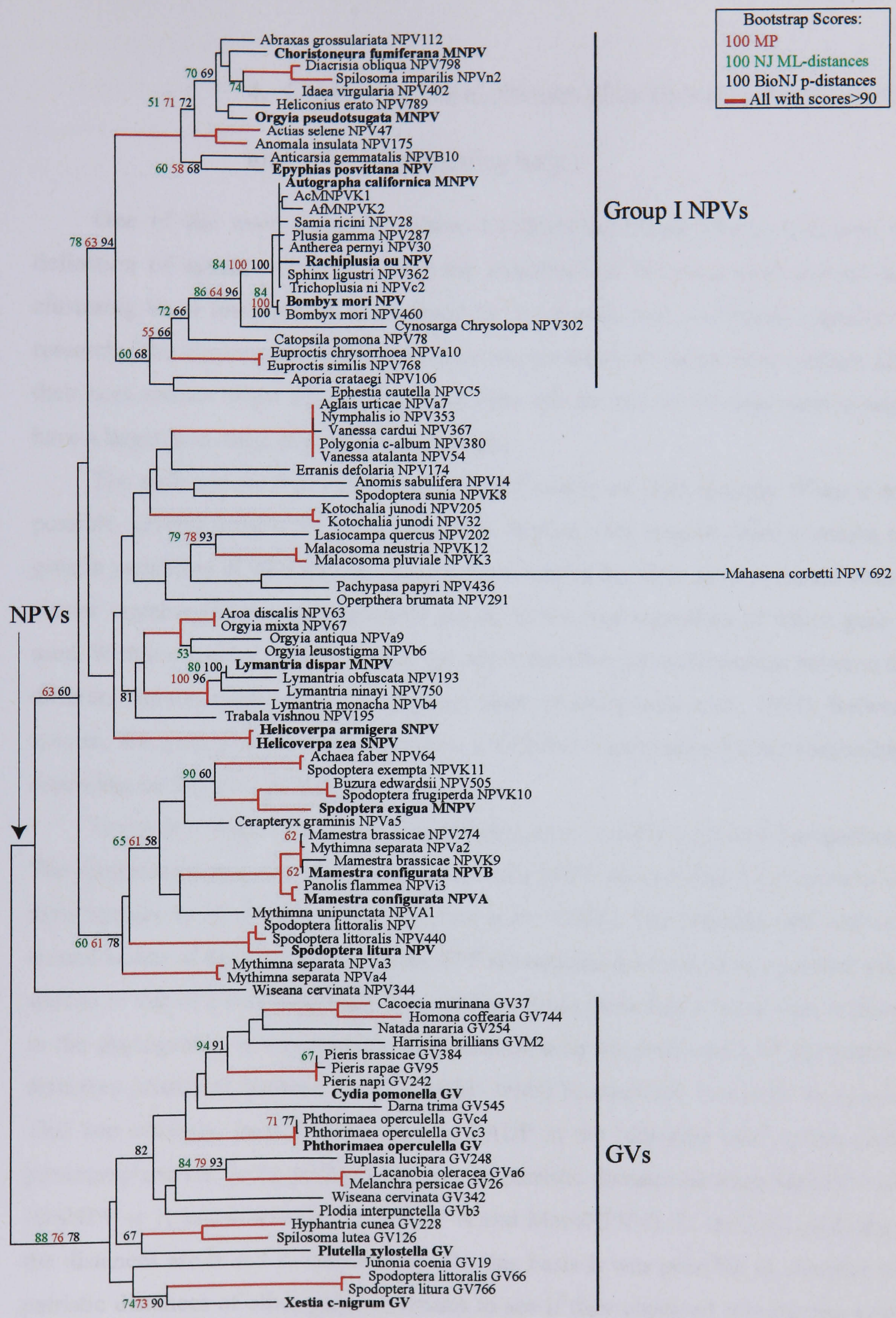


Figure 24: *Lef-8* Phylogeny. ML tree, 1000 boostrap replicates were performed for the MP, NJ with ML distances and BioNJ methods. **Orgyia pseudotsugata MNPV:** in bold sequences from complete genomes, Actias selene NPV47: in black sequences acquired in this study expect for Spodoptera littoralis NPV.

4. Clustering Viral Strains Into Species: Do the polh and lef-8 phylogenies help?

One of the main questions about lepidopteran baculoviruses concerns the definition of species. Understanding the relationships between viral isolates and clustering them into species units could be a valuable tool for future biopesticide research. The current trend of baculovirus nomenclature to name virus isolates after their host species might not reflect actual virus species and some virus species might have a larger host range than originally thought.

The *polh* and *lef-8* phylogenies are based mostly on field isolates. When it was possible, several isolates from the same host species were used in order to assess the genetic variability of viral species. Viruses representing the same species should always cluster together (form a monophyletic group) in the tree regardless of which gene is used. Within a species, recombination can occur therefore the relationships between the different isolates might differ for different genes (Koufopanou et al., 1997). Between species, the gene phylogenies might give a different hypothesis of interrelationships depending on the gene evolution.

Taxonomic hypotheses have also resulted from complete genome comparisons. The comparative analysis of the *Helicoverpa zea SNPV* showed that it belonged to the same species as *H. armigera SNPV* (Chen et al., 2002). The genomic analyses of a second isolate of *Mamestra configurata NPV* showed that it belonged to a separate virus species to that of a first isolate (Li et al., 2002a). Since these four viruses were included in the phylogenies, it was possible to determine both for *polh* and *lef-8* the patristic distances (distances between isolates within trees) between the two pairs of species. This was obtained from the ML trees in PAUP in the 'describe tree' option under parsimony criteria. In the *polh* phylogeny, the patristic distance between HaSNPV and HzSNPV is 1, and 4 between MacoNPV-A and MacoNPV-B. In the *lef-8* phylogeny, the distances are 0 and 2, respectively. On this basis it was possible to compare the patristic distances of other pairs of viruses to see if they clustered into species units. Pairs of virus isolates with patristic distances falling bellow or within the range defined above (1-4 for the *polh* tree, and 0-2 for the *lef-8* tree) were recorded.

However, the assumption behind this method was that all baculovirus species had similar rates of nucleotide substitution. If a species has high rates of substitutions

between strains, it might be split based on the comparisons to slower evolving species. To improve the inference, other parameters were employed. When it was possible, the clustering of viruses into species was based on both the *polh* and *lef-8* genes. Species should also be monophyletic. This allowed further discrimination in cases where the patristic distances conflicted with the tree topologies. This happened mostly when the distance was on the higher range of the cut-off for inclusion into a species, but also when the distance was within the range but the viruses did not form a monophyletic group.

a) Nucleopolyhedrovirus

The *polh* phylogeny comprised 114 NPV virus samples. Patristic distances were compared between each pair of isolates. A total of 67 NPV virus species were defined based on both *polh* and *lef-8* phylogenies. Of these species, 28 comprised more than one virus sample (figure 25).

(1) Are isolates from the same host species representative of the same virus species?

There have been a number of documented cases of host species from which different baculovirus species have been isolated. It is not surprising that this has been found mostly in well-studied forest and agricultural pests such as *Choristoneura fumiferana* (MNPV, NPV-DEF and GV) and *Trichoplusia ni.* (AcMNPV, SNPV and GV) (Blissard et al., 2000). So to fully explore the diversity of baculoviruses it was essential not to overlook isolates from host species already sampled in the phylogeny (from the database or in this study).

Nine samples, including BmNPV-460, were duplicates of viruses found in the database (figure 25). The fact that they clustered into species with the published sequences showed that there are well-defined virus species corresponding to a particular host. However, this was not the case for all the samples. The Orgyia pseudotsugata NPV-B5 sample belonged to neither of the 2 virus species (OpMNPV or OpSNPV) already isolated from this host (figure 25).

The taxonomy of the Plusinae sub-family of the Noctuidae is fairly confused. This gave rise to a number of names for what should probably be the same moth species. The species *orichalcea* has been classified within three genera: *Plusia, Trichoplusia* and *Thysanoplusia*. Viruses have been described under these three names. I have kept the name of these viruses as they were published and stored in various data banks. This

actually allowed a different name to be attributed to each of the three virus species derived from this host. The *polh* phylogeny also showed that a sequence published for Thysanoplusia orichalcea NPV (Cheng et al., 1998) had been submitted to gene bank as Plusia orichalcea NPV, hence contributing to further confusion (figure 25).

This work confirmed that given a wide enough sampling, a number of host species are susceptible to more than one species of NPVs. Nevertheless more often than not only one sample was available per host species so this might not be generalised. There are also examples where all the isolates from one host gave the same virus, such as *Operopthera brumata NPV* (figure 25), showing that there is a basis for naming baculoviruses after their host.

(2) Some virus species infect more than one host.

A baculovirus such as AcMNPV is known for its broad host range (Lacey et al., 2002, Payne, 1986). However, apart for *Lymantria dispar* NPV (Barber et al., 1993) and *Mamestra brassicae* NPV (Doyle et al., 1990), *Anagrapha falcifera* NPV (Hostetter & Puttler, 1991), *Euproctis chrysorrhoea* NPV (Cory et al., 2000), few baculoviruses have been studied as much as AcMNPV and therefore their host range remains largely undetermined. This phylogenetic study clusters into species, virus isolates from various sources and hosts, representing a valuable means for exploring the host range of some virus species.

Some virus species seem confined to a specific Lepidoptera genus. This was the case for two viruses of lymantriid hosts. *Orgyia pseudotsugata SNPV* was found in moth species of the genus *Orgyia*: *O. pseudotsugata* and *O. antiqua.* These two viruses (OpSNPV and O antiqua SNPV) have previously been found to share relatively similar restriction endonuclease profiles and OpSNPV was found to readily infect larvae of *O. antiqua* (Richards et al., 1999). This supports the phylogenetic clustering of these two viruses into the same species. *Lymantria dispar MNPV* was found in moth species of the genus *Lymantria*: *L. dispar, L. obfuscata* and *L. ninayi* (figure 25). The high degree of host specificity of LdMNPV has been tested against 46 non-target lepidopteran species, which all proved non-permissive (Barber et al., 1993). *Malacosoma californicum NPV* was also found to have infected larvae of the lasiocampiid *Malacosoma californicum pluviale* and *M. neustria*.

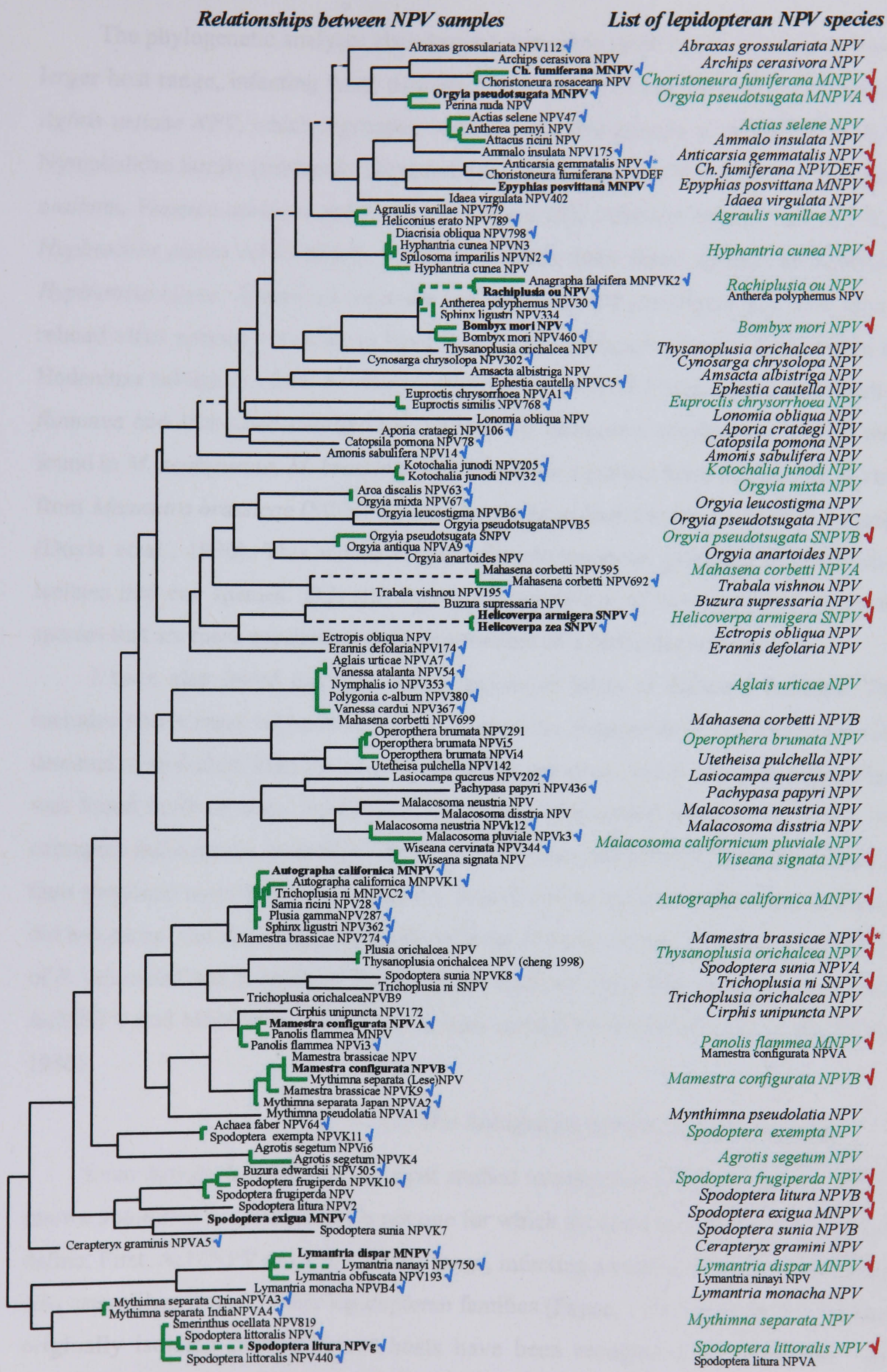


Figure 25: List of lepidopteran *Nucleopolyhedrovirus* species. See figure 26 for legend; * could represent a different species than the one proposed before, in regular script: isolates of disputable status; * a different isolate was used in the lef-8 phylogeny but the relationships with other viruses are the same and thus treated as the same species.

The phylogenetic analyses also showed that some virus species seemed to have a larger host range, infecting hosts of particular families of Lepidoptera. This included *Aglais urticae NPV,* which regrouped isolates from five species of butterflies from the Nymphalidae family (subfamily Nymphalinae), *Aglais urticae, Nymphalis io, Polygonia c-album, Vanessa atalanta* and *V. cardui* (figure 25). Another notable case is that of *Hyphantria cunea NPV*, which included isolates from three species of Arctiidae, *Hyphantria cunea, Spilosoma imparilis* and *Diacrisia obliqua* (figure 25). Two closely related virus species appeared to have broad and overlapping host ranges within the Hadeninae subfamily of the Noctuidae. *Panolis flammea NPV* was infectious to both *P. flammea* and *Mamestra configurata.* Isolates of *Mamestra configurata NPV-B* were found in *M. configurata, M. brassicae* and *Mythimna separata* hosts (figure 25). A virus from *Mamestra brassicae* (MbNPV) has already been found to have a broad host range (Doyle et al., 1990). This would support the phylogenetic grouping of these virus isolates into one species. This still leaves the possibility of having strains within the species that are more specialised and more virulent to a particular host species.

I have also found the same virus species in hosts of different families. This includes two viruses of noctuid hosts: *Spodoptera frugiperda NPV,* which was also detected in an isolate from the geometrid *Buzura edwarsii,* and *S. littoralis NPV*, which was found in *Smerinthus ocellata* of the family Sphingidae (figure 25). These two examples indicate the possibility for viruses to cross infect hosts distantly related to their proposed main hosts. However, the lack of ecological data from these two cases did not permit the assessment of the prevalence of these viruses in natural populations of *B. edwardsii* and *S. ocellata*. However this does not seem impossible as viruses like AcMNPV and MbNPV are infectious to non-noctuid hosts (Doyle et al., 1990, Payne, 1986)

(3) The Autographa californica MNPV species

Even though AcMNPV is the most studied baculovirus in the laboratory, little is known about it in the field and it is not one for which the species boundaries are easy to define. First, AcMNPV has a broad host range, infecting a number of noctuid larvae but also caterpillars from 13 other lepidopteran families (Payne, 1986). A number of viruses originally isolated from different hosts have been recognised as "Acal-like" and therefore classified within the AcMNPV species (Black et al., 1997). They include Anagrapha falcifera NPV, Rachiplusia ou NPV, Galleria mellonella MNPV, Spodoptera exempta MNPV and Trichopluisa ni MNPV (Blissard et al., 2000).

A number of isolates used in this study turned out to have sequences identical or very similar to that of the sequence from the AcMNPV genome. This was not excessively surprising as the samples came from the NERC-CEH collection, an institute where numerous bioassays experiments have been conducted over the years. As most collecting details of the samples have been lost there was no way to check back on the origins of the samples, but it is possible that samples such as Samia ricini NPV-28 (SariNPV-28) and Sphinx ligustri NPV-362 (SphinxlNPV-362) might have been the results of host range determination assays of AcMNPV.

The *lef-8* phylogeny showed the AcMNPV viruses closely related to BmNPV (figlef-8tree). This position was consistent with the complete genome phylogenies (chapter 2). The patristic distances (value of 0) for this gene also suggested that Rachiplusia ou NPV (RoNPV), Anagrapha falcifera NPV-k1 (AfMNPV-k1) and Antherea polyphemus NPV-30 (AnpoNPV-30) were also part of the AcMNPV species. However, the *polh* phylogeny gave a completely different view of the relationships between these viruses. The tight AcMNPV group of the *lef-8* tree was split in the *polh* tree. RoNPV, AfMNPV-k1, AnpoNPV-30 and Sphinx ligustri NPV-334 were part of a group closely related to BmNPV, in accordance with the lef-8 phylogeny. AcMNPV, AcMNPV-k2, TnMNPV-c2, SariNPV-28, SphinxlNPV-362 and Plusia gamma NPV-287 also grouped tightly together, but in a completely different position in the tree (figure 25). This suggested that the AcMNPV species might be split into AcMNPV and RoNPV based on the origin of the *polh* gene present in the virus genomes. RoNPV, which seems to have most of its genome closely related to that of BmNPV, might have been at the origin of AcMNPV, as they share high genomic similarity (Harrison & Bonning, 2002). A recombination event in the *polh* region seems a key event that defined the AcMNPV species.

The *polh* phylogeny (figure 23) provided some clues to the origin of the AcMNPV *polh* gene. It was most closely related to the *polh* of Mamestra brassicae NPV-274 (MbNPV-274). However this virus in the *lef-8* phylogeny seemed to be of the same background as *Mamestra configurata NPV-B*. This might also highlight another case of recombination. Although the topology in this part of the tree was poorly supported, the best *polh* tree suggested that the AcMNPV *polh* gene was most closely related to that of virus species from Plusinae hosts and more widely noctuid hosts. *Trichoplusia ni* is permissive to AcMNPV, MbNPV and RoNPV (Vail et al., 1999). Cross infections by different virus species of hosts with multiple susceptibility, such as

T. ni, could have been at the origin of the recombination event that led AcMNPV to harbour a *polh* gene of a different evolutionary background than the rest of its genome. The promoter region of AcMNPV is identical to that of BmNPV (Zanotto et al., 1993), suggesting that any recombination in the *polyhedrin* gene of AcMNPV would have had to take place downstream of the promoter. The AcMNPV polh promoter has been shown to lead to 30% more gene expression than that of MbNPV (Cameron & Possee, 1989). The AcMNPV promoter would therefore provide a greater selective advantage than that of MbNPV or closely related viruses. Its replacement by a promoter giving lower gene expression would probably be selected against and thus not retained in the evolution of AcMNPV.

The replacement of the *polh* gene in AcMNPV must have provided the virus with a great selective advantage to be maintained in the genome. It might have been a factor of the extended host range observed for AcMNPV, which is different to that of RoNPV (Vail et al., 1999).

This study of AcMNPV and closely related species suggests a number of taxonomic changes raising the strains Rachiplusia ou MNPV and Spodoptera exempta NPV to the rank of species and incorporating new isolates.

Baculovirus Species

- *Autographa californica MNPV*

  Strains and isolates used in this study

  - Autographa californica MNPV-P.Vail
  - Trichoplusia ni MNPV-c2
  - Plusia gamma NPV-287
  - Samia ricini NPV-28
  - Sphinx ligustri NPV-362

  Other Strain from the ICTV list not included in this study

  - Galleria mellonella MNPV

- *Rachiplusia ou MNPV*

  - Anagrapha falcifera MNPV-P.Vail
  - Antherea polyphemus NPV-30 (could be raised to species)
  - Sphinx ligustri NPV-334

- *Spodoptera exempta NPV*

  - Spodoptera exempta NPV-K11
  - Achaea faber NPV-64

b) *Granulovirus*

The *polh* phylogeny comprised 34 granulovirus samples. Patristic distances were compared between each pair of isolates. A total of 25 GV virus species were defined based on both *polh* and *lef-8* phylogenies. Of these species, only six comprised more than one virus sample (figure 26).

Harrisina brillians GV, Homona coffearia GV and *Phthorimaea operculella GV* were present in duplicate isolates from the same host species, whereas *Choristoneura fumiferana GV, Spodoptera littoralis GV* and *Pieris brassicae GV* included isolates from different host species. *Spodoptera littoralis GV* and *Pieris brassicae GV* were regrouping isolates from closely related hosts species of the same genera. *Choristoneura fumiferana GV* was found in hosts of the family Tortricidae, *C. furamifera* and *Cacoecia murinana.*

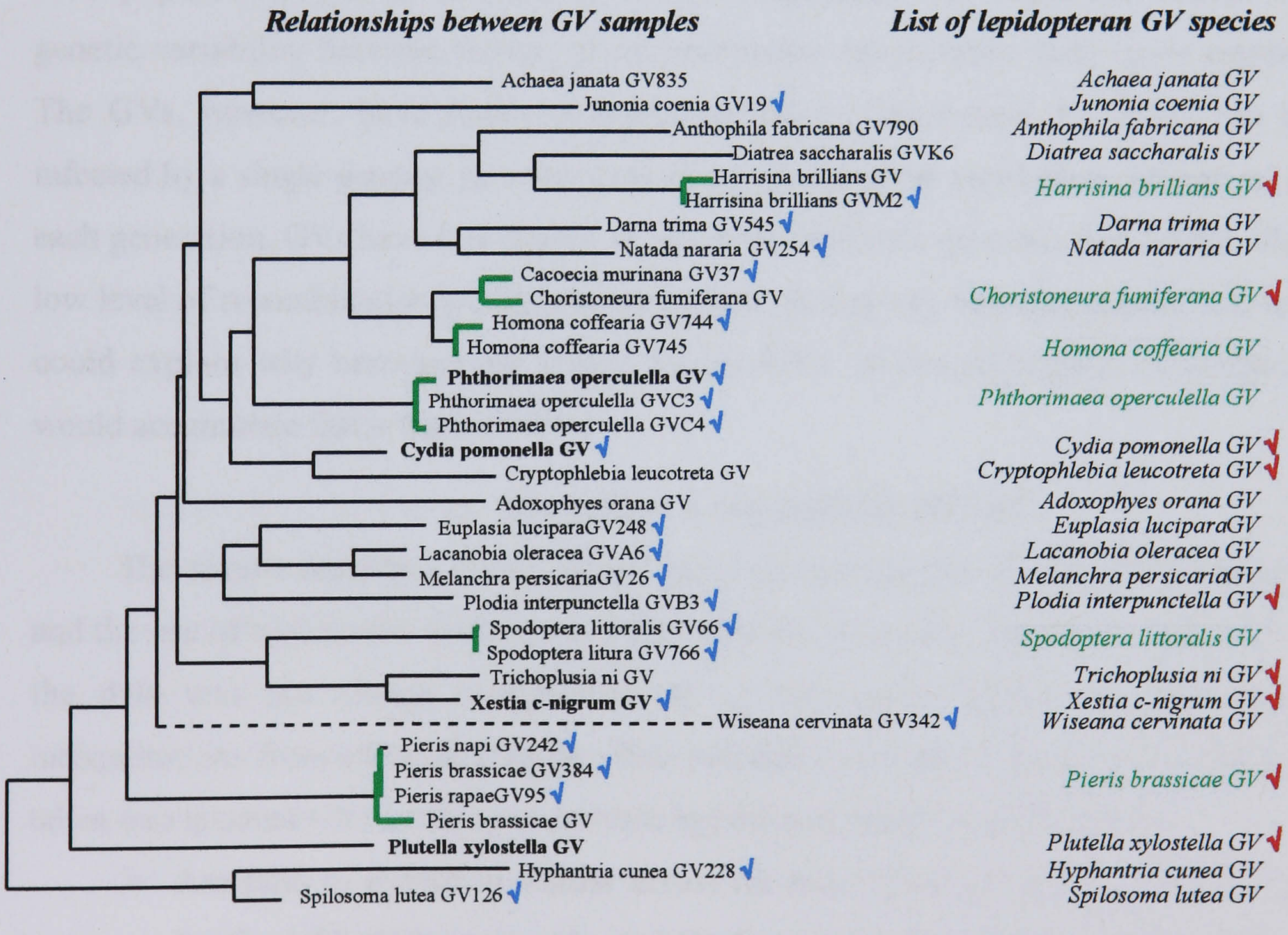


Figure 26: List of *Granulovirus* species. In *italics:* virus species, in green: species including several isolates, √ species proposed in the last ICTV report. The relationships between viral isolates are based on the polh phylogeny; √ isolates used in lef-8 phylogeny; in green: branches showing isolates clustering into the same species; - - dashed lines represent branches shortened for the figure.

Of note was a difference observed in the patristic distances between GVs and NPVs for the *lef-8* phylogeny. When distances observed between GV isolates of the same species were similar to those of NPVs in the polh phylogeny, they sometimes

appeared high or borderline for the *lef-8* tree. This was the case for the isolates Spodoptera littoralis GV-66 and S. litura GV-766 with a distance of 2 in the *lef-8* tree, equal to the distance between the two NPV species isolated from *Mamestra configurata*. The null distance in the *polh* tree made it clear that these two isolates belonged to the same GV species. However, the initial conflict residing in the comparison of patristic distances reflected on the differences that exist between GVs and NPVs including that they might not evolve in the same fashion.

Besides the patristic distances, the *lef-8* phylogeny (figure 24) also showed that overall the GVs have longer branches than the NPVs. The difference in branch lengths between NPV and GV species could be related to the biology of these two groups of baculoviruses. NPV occlusion bodies comprise multiple virus particles. This implies that the hosts are always infected by several genomes and that recombination can happen frequently. Recombination would promote and maintain a variable gene pool for NPV populations, just as sex does in cellular organisms, but would also reduce the genetic variability between strains, as recombination would make them more similar. The GVs, however, have singly occluded particles. This means that hosts can be infected by a single genome or a few genomes within a given population. Therefore, at each generation, GVs have less chance to recombine between genomes than NPVs. This low level of recombination would promote genetic divergence between populations and could explain why branches are longer for the GVs in the phylogeny, as mutation would accumulate faster than in NPVs.

c) Discussion: Using patristic distances

The results described above undoubtedly showed the usefulness of phylogenies and the use of within tree distances to define clusters of isolates. However, interpreting the data was not always straightforward. A few cases highlighted conflicting interpretations from either phylogeny. This outlined a number of issues that should be taken into account when making taxonomic hypotheses based on phylogenies.

- Are rates of evolution similar across the tree? It did not seem to be the case for the *lef-8* phylogeny with patristic distances within species being higher in the GVs than in the NPVs. However, within the *polh* tree, the distances seemed consistent across the tree. This gene might therefore remain an appropriate marker to identify baculovirus species.
- Is it appropriate to make taxonomic hypotheses based on a single gene? Again the *lef-8* phylogeny alone did not permit the definition of some of the

baculovirus species. This can probably be attributed to lack of variation of this gene fragment at the species level (0-2). The *polh* phylogeny showed more variation at the species level (0-4) and discrimination was therefore easier. The case of the AcMNPV isolates showed that *lef-8* on its own was not enough to make the discrimination, as this species appeared as a recombinant.

The phylogenetic species concept is invaluable for defining species of baculoviruses. The analyses of patristic distances by clustering isolates into virus species proved useful for assessing potential host range. No previous baculovirus phylogenetic analyses have comprised as many isolates from so many hosts. The lack of suitable gene sequences for many isolates often made it difficult to compare them. Sequencing complete genomes is the only way to completely assess the genetic differences or similarity between baculoviruses isolates, however this is still currently a costly process and impractical for many baculoviruses from non-economically important hosts. The genetic bases for the difference in biology between baculovirus strains remain to be discovered. This would be crucial to understand the making up of the species and the speciation process in baculoviruses.

5. Measuring the incongruence between the polh and lef-8 phylogenies

Examination of the *polh* and *lef-8* phylogenies quickly revealed conflicting topologies, particularly in the positioning of the AcMNPV species. This was attributed above to a recombination event in the *polh* region of AcMNPV. The congruence of both phylogenies remains to be tested to assess if there is a significant difference between the tree topologies when AcMNPV is excluded, or if recombination was rampant during the evolution of baculoviruses.

The tests of congruence between the *polh* and *lef-8* data sets were performed on reduced matrices which included one sample of each of the baculoviruses species detected above, which are in common to both sets, and excluded AcMNPV. This amounted to 64 taxa. The phylogenetic analyses were conducted as before.

a) KH and SH tests

The *polh* and *lef-8* trees obtained were examined against each data set using the KH (one-tail) and SH tests. As before the trees did not pass the tests for the other data set ($p<0.001$). So each gene was underlining a different hypothesis of baculovirus

evolution. However, it was not possible to discriminate with these tests as to whether the differences in tree topologies were due to genuine gene evolution differences or to poor signal and resolution of the phylogenies. A strict consensus between the two trees showed that there were congruent parts of the topologies, as 13 nodes were exactly conserved.

b) TreeMap evaluation

To explore in more detail the congruence between both *polh* and *lef-8* phylogenies, the trees were analysed using the program TreeMap (Page, 1995). This program was designed to study co-evolution between hosts and parasites. It implements a graphic representation of the associations between the species of both trees and calculates the number of nodes showing co-speciation events.

The tanglegram of the *polh* and *lef-8* trees (figure 27) showed that most of the disagreement between both phylogenies was between species of the group II NPVs. A heuristic reconstruction to reconcile both trees suggested 30 possible events of co-speciation between both genes, corresponding to 46.8% of the nodes. This represented a much higher number than found by the strict consensus because TreeMap allowed for gene duplication and loss in the reconstruction thus increasing the number of co-speciation events between both genes.

c) Bootstrap

To assess whether the parts of the trees where no co-speciation between both genes was observed could be due to poor phylogenetic resolution, bootstrap tests were performed on both phylogenies. The nodes supported by bootstrap scores superior to 50 were mapped on the tanglegram (figure 27). This showed that there was a clear correspondence between the parts of the trees well supported by bootstrap and those nodes, which showed co-speciation. However there were a few exceptions, such as PxGV, which possibly represent baculoviruses which had undergone recombination events in the past.

Most of the conflict between both *polh* and *lef-8* tree topologies resided in poorly supported nodes. These included most of the group II NPVs and relationships within the group I NPVs and GVs. This highlighted further the weaknesses found in both phylogenies. The strengths found in both *polh* and *lef-8* phylogenies corresponded to the relationships of viruses of the same species.

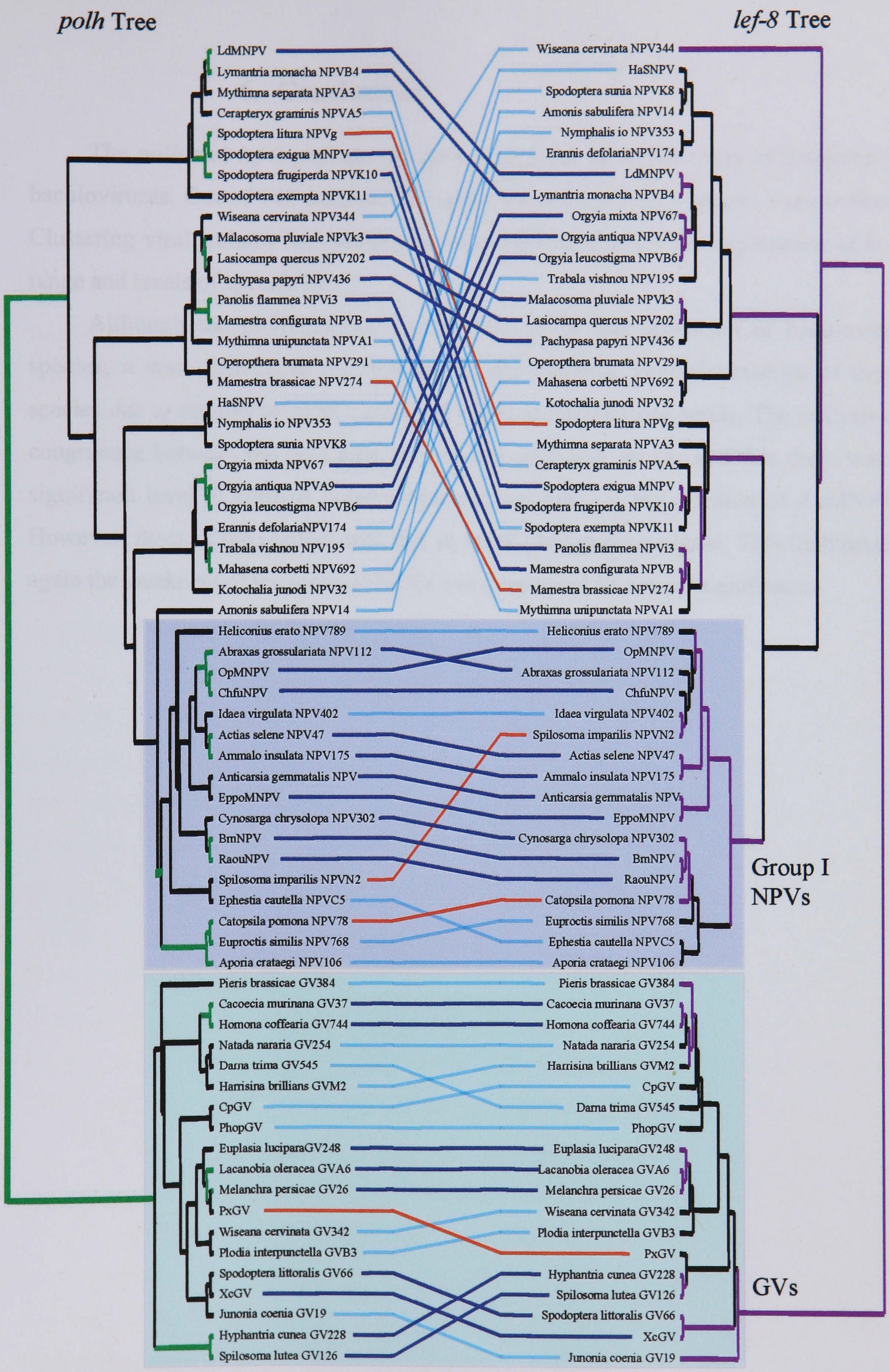


Figure 27: Congruence between the *polh* and *lef-8* trees. On the trees, coloured branches have a bootstrap support >50. Navy blue lines in the tanglegram link well supported nodes agreeing in both trees; orange lines correspond to viruses, which have conflicting,well supported positions. Within the coloured boxes are the well supported clades: group I NPVs and GVs.

6. Conclusions

The *polh* and *lef-8* phylogenies gave a measure of the diversity of lepidopteran baculoviruses. Out of 148 isolates, 92 lepidopteran baculovirus species were defined. Clustering viral isolates into baculovirus species also enabled the exploration of host range and levels of specificity.

Although the phylogenetic signal allowed for the definition of baculovirus species, it was difficult to reconstruct the phylogenetic interrelationships of those species due to the low level of support of nodes at intermediate levels. The analysis of congruence between the *polh* data set and the *lef-8* data set showed that there was a significant level of conflict between the phylogenies, i.e. the position of AcMNPV. However, most of the conflict was due to weak phylogenetic signal. This highlighted again the weakness of the group II NPVs and questioned its natural significance.

E. Lepidopteran Hosts Shape Baculovirus Evolution

1. Introduction: Bases of Baculovirus-Host Co-Evolution.

When individuals from different species interact, the associations can range from mutualism to parasitism. In the case of a host-pathogen interaction, such as lepidopteran larvae and baculoviruses, the antagonistic relationships between both species imply that they apply reciprocal selection pressure to each other. According to the Red Queen theory (Van Valen, 1973), pathogens select for fitter hosts and the hosts select for the pathogens that exploit them best (Combes, 2001). This arms race between host and pathogens is believed to lead to the specialisation of interspecific associations (Ehrlich & Raven, 1964). This is thought to account for much of the evolutionary history of the diversification of life (Thompson, 2002).

Baculoviruses, like any virus, have some of the most intimate associations with their hosts as they reproduce in their cells. Caterpillars have several lines of defence against baculovirus infection. The peritrophic membrane of the insect midgut has been found to provide some protection against baculovirus infection (Wang & Granados, 2000). The next level of resistance is within the cells. Some larvae have been found to slough the infected midgut epithelial cells thus precluding the spread of infection (Washburn et al., 2001). Hemocytes have also been found to remove viruses from the haemolymph therefore participating in the reduction of viral load and subsequent fatal infection (Trudeau et al., 2001). Lastly a number of genes have been found to expand the host range of baculoviruses (Argaud et al., 1998, Chen et al., 1998, Chen & Thiem, 1997), emphasizing the intimacy of the interaction between a baculovirus and its host.

The complexity of the interactions between baculoviruses and their hosts suggests that the evolution of their associations should lead to the specialisation of the viruses. This process would reduce the number of host species that are suitable to a particular virus species. The phylogenetic constraint of host specialisation leaves fewer possibilities for conquering new host species over evolutionary time (Price, 2002). Hence a baculovirus species infecting a new host species is more likely to be able to do so in species closely related to its current host. In this case, closely related viruses should infect more closely related hosts. The phylogenetic pattern of host use in baculoviruses should therefore reflect the constraint imposed by specialisation.

The main aim in this section is to determine whether and to what extent hosts have influenced baculovirus evolution. Several steps will be undertaken to address this question. (i.) Is there any evidence of phylogenetic conservatism of host affiliations? (ii.) Is it possible to infer the family of the ancestral lepidopteran baculovirus host?

2. Methods: Mapping Baculovirus Host Use

a) Baculovirus Phylogenies

The *polh* and *lef-8* phylogenies obtained in the previous section were used for this study. Both genes present alternative hypotheses of baculovirus evolution. The *polh* conflicts more with the complete genome phylogeny but it represents a much larger sampling of virus species than the *lef-8* tree. The *lef-8* phylogeny, as well as the topologies obtained from the congruence study, will be used as a point of comparison for the results from analyses based on the *polh* phylogeny. Baculovirus species topologies were derived from trimming the trees so that each species is represented by one branch. Equivocal strains, such as *Antherea polyphemus NPV* and *Spodoptera litura NPV-g* were treated as separate species.

b) Evolution of the Lepidoptera

There are no phylogenies of the Lepidoptera that encompass the relationships between the hosts of all the baculoviruses used in this study. Although it is subject to debate, the traditional morphological classification at the family level was the best option I could find for understanding the relationships between the moths and butterflies linked to this study. For consistency, I used the classification of families and superfamilies of Scoble, as shown in table 15 (Scoble, 1992).

Within the superfamily Noctuoidea, recent morphological revisions (Kitching & Rowlins, 1999) and phylogenetic studies (Mitchell et al., 2000) have shown that the Noctuidae are paraphyletic (split into several lineages). The authors however did not propose that the family Noctuidae should be divided at the moment, as the relationships of the main subfamilies have yet to be completely resolved.

Because there were no trees available for the Lepidoptera, it was not possible to look for events of co-speciation between the baculoviruses and their hosts at species level. Nevertheless, it is possible to assess the degree of phylogenetic conservatism based on higher taxonomic levels, such as family and superfamilies.

table 15: Classification of the Lepidopteran Hosts of the Baculoviruses used in this study – Higher level classification of the Lepidoptera after Scoble (1992).

Superfamilies	Families	Genera
EXPORIA		
Hepialiodea	Hepialidae	*Wiseana*
HETERONEURA		
Tinoidea	Psychidae	*Kotochalia, Mahasena*
Yponomeutoidea	Plutellidae	*Plutella*
	Glyphipterigidae	*Anthophila*
Gelechioidea	Gelechiidae	*Phthorimaea*
Tortricoidea	Tortricidae	*Adoxophyes, Archips, Choristoneura, Cydia, Cryptophlebia, Epiphyas, Homona*
Zygaenoidea	Zygaenidae	*Harrisina*
	Limacodidae	*Darna, Natada*
Pyraloidea	Pyralidae	*Diatrea, Ephestia, Plodia*
Geometroidea	Geometridae	*Abraxas, Buzura, Ectropis, Erannis, Idaea, Operopthera*
Papilionoidea	Pieridae	*Aporia, Pieris, Catopsila*
	Nymphalidae	*Aglais, Agraulis, Heliconius, Junonia, Inachis, Polygonia, Vanessa*
Bombycoidea	Lasiocampidae	*Lasiocampa, Malacosoma, Pachypasa, Trabala*
	Saturniidae	*Actias, Antherea, Attacus, Lomonia, Samia*
	Bombycidae	*Bombyx*
	Sphingidae	*Sphinx, Smerinthus*
Noctuoidea	Notodontidae	*Cynosarga*
	Lymantriidae	*Aroa, Euproctis, Lymantria, Orgyia, Perina*
	Arctiidae	*Ammalo, Amsacta, Diacrisia, Hyphantria, Spilosoma, Utetheisa*
	Noctuidae	*Achaea, Autographa, Anagrapha, Agrotis, Amonis, Anticarsia, Cerapteryx, Cirphis, Euplasia, Heliothis, Helicoverpa, Lacanobia, Mamestra, Melanchra, Mythimna, Panolis, Plusia, Rachiplusia, Spodoptera, Thysanoplusia, Trichoplusia, Xestia*

c) Assessing the phylogenetic conservatism of lepidopteran host use

The 94 baculovirus species of the *polh* tree were isolated from 111 different host species spanning 20 families in 11 superfamilies of Lepidoptera. The hosts of 64 virus species from the *lef-8* phylogeny also represent the 11 superfamilies. The families and superfamilies of the lepidopteran hosts were mapped on the *polh* and *lef-8* baculovirus phylogenies using parsimony approaches in MacClade 4 (Maddison & Maddison, 2000).

To investigate to what extent baculovirus evolution has been constrained by the relationships between their hosts, a number of statistical tests were performed. The

lepidopteran families were coded as separate binary characters, where 1 is for virus species isolated from hosts of the family considered. The data matrix comprised 20 characters, one per lepidopteran family. Baculoviruses infecting hosts of a single family have only one character coded as 1, whereas baculoviruses found in hosts from different families, such as AcMNPV, have several characters coded 1. The phylogenetic data (figures 25-26) was used to ascertain the host range.

To assess if the pattern of lepidopteran host use observed from the baculovirus phylogenies could have been obtained randomly, the trees and the family binary dataset were randomised 10,000 times. The lengths of the trees (number of steps) obtained from random trees or random datasets give a null random distribution, against which the observed data can be compared (length of the 'true tree'). The PTP test (permutation tail probability) in PAUP was used to randomise the data set, as in similar studies (Lopez-Vaamonde et al., 2003, Wahlberg, 2001). The tree randomisation (Huelsenbeck, 1991) was performed in MacClade.

d) Searching for the Ancestral Lepidopteran Baculovirus Host

A different approach was used to encode the characters in order to reconstruct the ancestral host infected by baculoviruses. Similar to studies in insect host plant use (Lopez-Vaamonde et al., 2003), the families and superfamilies of Lepidoptera were coded as unordered multistate characters. The family character had 20 different states and the superfamily character 11 states. Most virus species only infect caterpillars from one lepidopteran family. However AcMNPV was found in the phylogenies in isolates from insects belonging to the families Saturniidae, Sphingidae and Noctuidae, which are from the Bombycoidea and Noctuoidea superfamilies. Similarly *Antherea polyphemus NPV*, *Spodoptera frugiperda NPV* and *S. littoralis NPV* also infect hosts from different families. For all these four baculovirus species, the characters were treated as polymorphic, stating that they infect hosts from various families. Phylogenetic data (fig) was used to ascertain virus host range. This was complemented by host range data available for AcMNPV, RaouNPV (*Anagrapha falcifera* NPV), MacoNPV-B (*Mamestra Brassicae* NPV) (Doyle et al., 1990, Hostetter & Puttler, 1991, Payne, 1986). This extended the host range of AcMNPV and RaouNPV to nine lepidopteran families out of the 20 considered and the host range of MacoNPV-B to five lepidopteran families.

The optimisation of the ancestral lepidopteran host on the baculovirus tree was performed in MacClade. The program utilises maximum parsimony algorithms to reconstruct character state in ancestral lineages based on the characters observed in extant taxa. The model of reconstruction ACCTRAN was favoured for representation purposes as it presents accelerated transformations of characters and therefore shows more readily host shifts in cases where alternative reconstructions are possible. The model DELTRAN is the more conservative alternative model, which preferably retains ancestral characters. All most parsimonious reconstructions (MPR), including ACCTRAN and DELTRAN, were considered to infer the family and superfamilies of the ancestral lepidopteran baculovirus host. The optimisation of host families and superfamilies on the virus phylogenies also allows the investigation of major transitions of host use that have occurred during the evolution of baculoviruses.

3. Results and Discussion

a) Phylogenetic conservatism of lepidopteran host use

Mapping the Lepidoptera Families on the baculovirus species phylogenies showed that more closely related viruses appear to infect more closely related hosts. This was particularly evident in some cases. For instance, five viruses isolated from caterpillars of the family Lasiocampidae are place in a monophyletic clade (figure 28). To assess how likely it is to obtain the observed degree of phylogenetic conservatism by chance, randomisation tests of the data (PTP) and the tree were performed.

The randomisation tests produced 10,000 replicates of the binary family data sets, which were mapped on the observed *polh* and *lef-8* baculovirus phylogenies, or 10,000 random trees onto which were mapped the observed binary family data set. The length (number of steps) of trees generated by all the replicates were plotted and compared to those observed (figure 29). The results of these tests showed that the probability of randomly obtaining the observed distributions of baculovirus host use is lower than 0.0001 as none of the 10,000 replicates reproduced the data observed. These are highly significant results attesting that the patterns of host use observed in all four baculovirus phylogenetic trees cannot be explained by chance. In other words, there is a correlation between the baculovirus phylogenies and the relationships of their hosts.

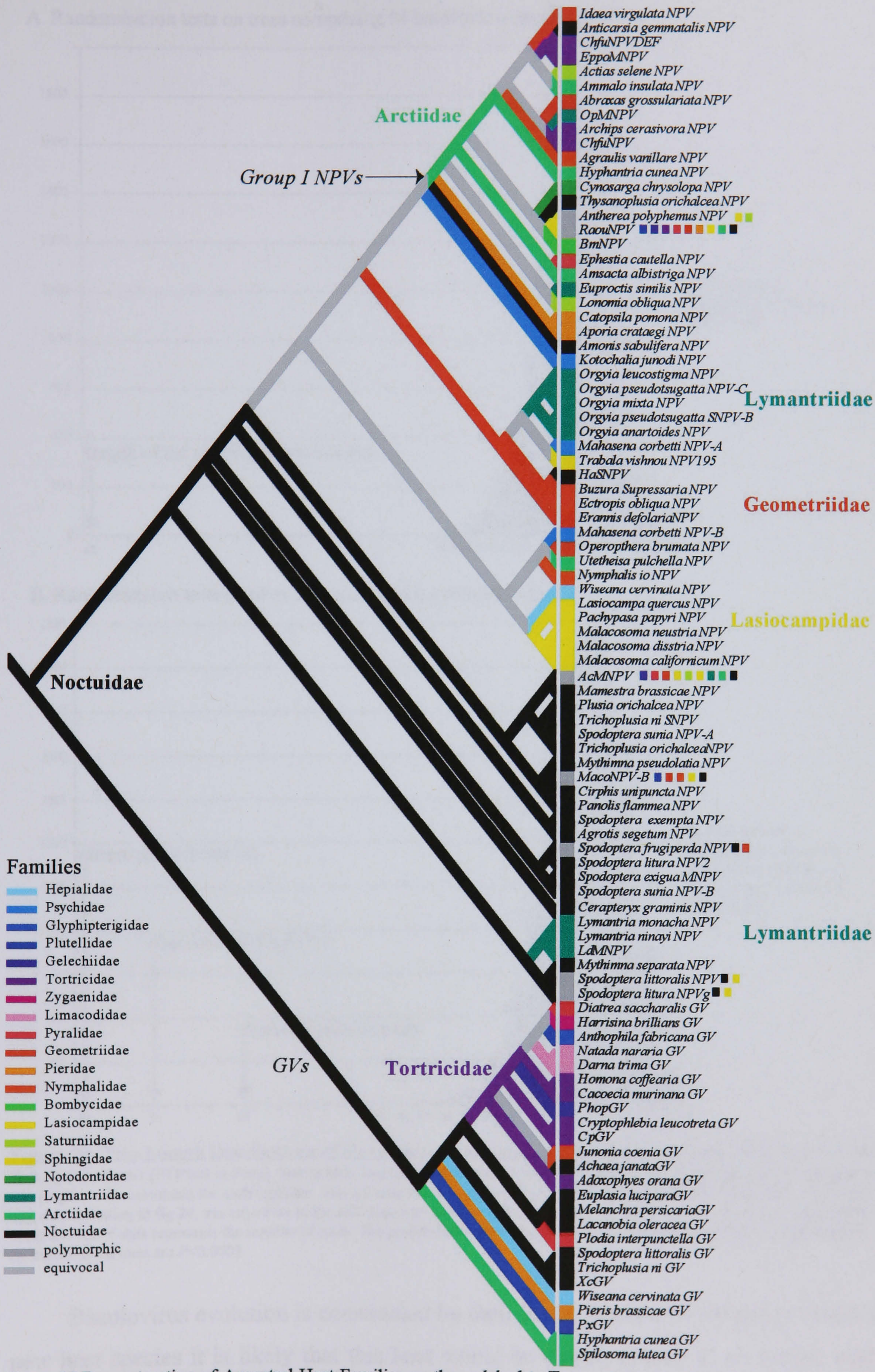
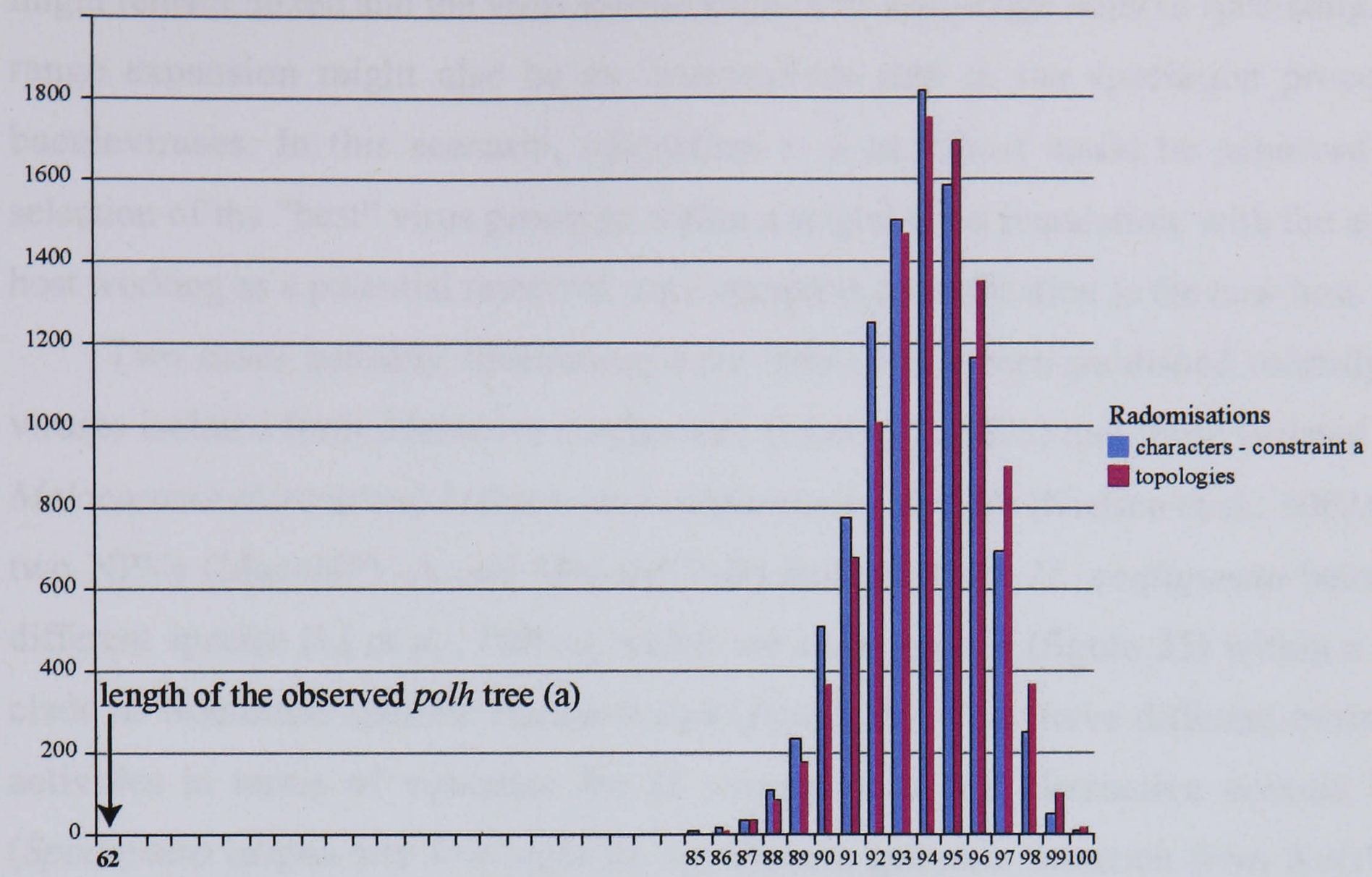


Figure 28:Reconstruction of Ancestral Host Families on the *polyhedrin* Tree. AcMNPV, Autographa californica MNPV; Bm-NPV, Bombyx mori NPV; ChfuMNPV, Choristoneura furamifera MNPV; ChfuNPVDEF, Ch. furamifera NPVDEF; EppoMNPV, Epyphias posvitta-naMNPV; HaSNPV, Helicoverpa armigera SNPV; LdMNPV, Lymantria dispar MNPV; MacoNPV-B Mamestra configurata NPV-B; OpMNPV, Or-gyia pseudotsugata MNPV; RaouNPV, Rachiplusia ou NPV; CpGV, Cydia pomonella GV; PhopGV, Phthorimaea operculella GV; PxGV, Plutella xylostella GV; XcGV, Xestia c-nigrim GV. Squares, next to the species names, indicate the characters states of polymorphic taxa.

A. Randomisation tests on trees comprising 94 baculovirus species



B. Randomisation tests on trees comprising 68 baculovirus species

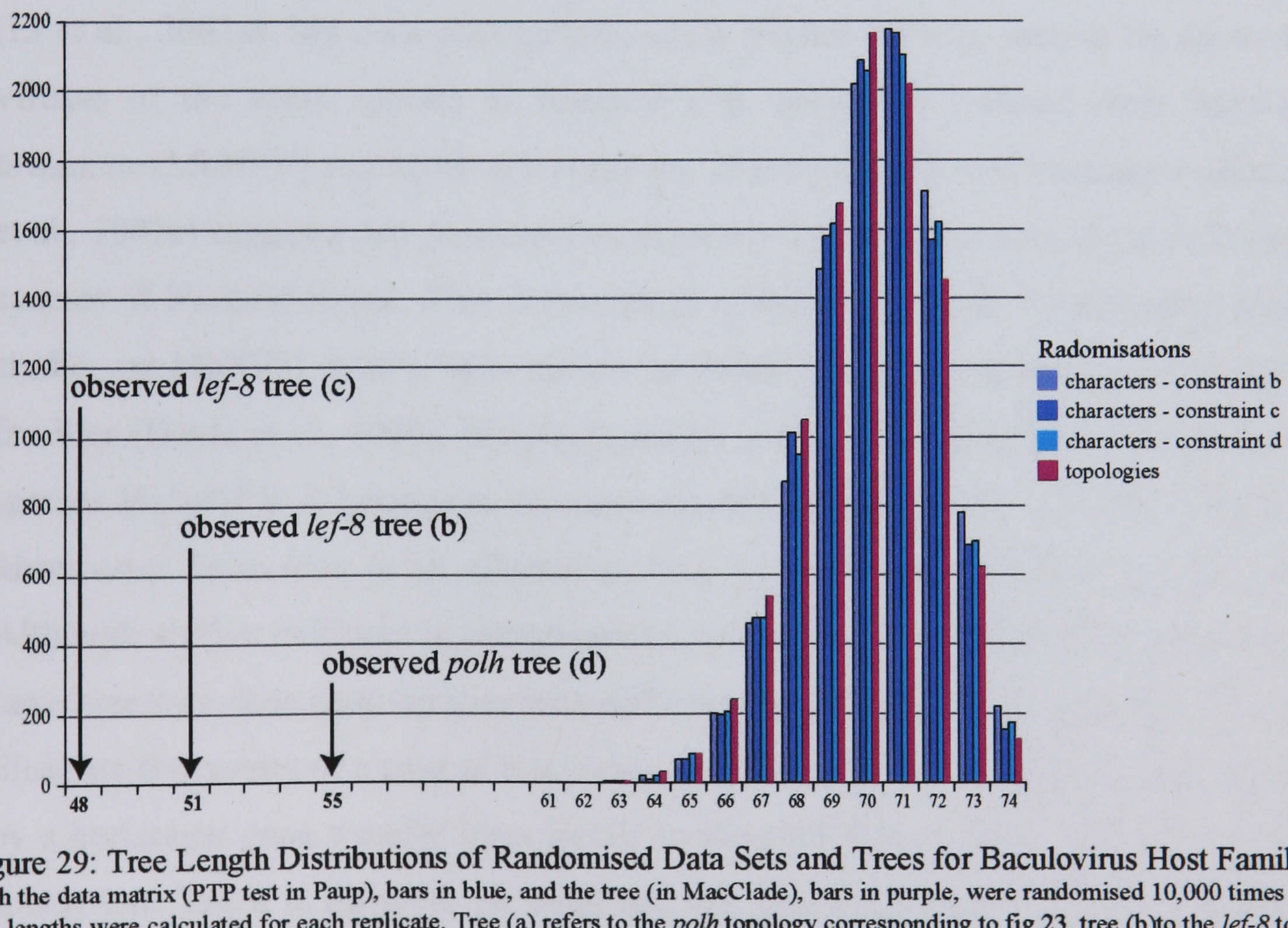


Figure 29: Tree Length Distributions of Randomised Data Sets and Trees for Baculovirus Host Families. Both the data matrix (PTP test in Paup), bars in blue, and the tree (in MacClade), bars in purple, were randomised 10,000 times and tree lengths were calculated for each replicate. Tree (a) refers to the *polh* topology corresponding to fig 23, tree (b)to the *lef-8* topology corresponding to fig 24, tree (c) refers to the *lef-8* topology corresponding to fig 27, tree (d)to the *polh* topology corresponding to fig 27. The Y axis represents the number of trees. The probabilities of obtaining the same character distribution of host family as on the observed trees are $P<0.0001$.

Baculovirus evolution is constrained by their hosts. When a baculovirus invades a new host species it is likely that this host would be closely related to its current host. Adaptation to the new host could lead to the isolation of specialised genotypes and speciation. In cases where both hosts share the same environment, the virus population

might remain mixed and the virus species expand its host range without speciating. Host range expansion might also be an intermediate step in the speciation process of baculoviruses. In this scenario, adaptation to a new host could be achieved after selection of the "best" virus genotype within a mixed virus population, with the ancient host working as a potential reservoir, until complete specialisation to the new host.

Two cases possibly illustrating these issues have been published recently, the viruses isolated from *Mamestra configurata* (Li et al., 2002a) and those isolated from *Malacosoma disstria* and *Malacosoma californicum pluviale* (Nielsen et al., 2002). The two NPVs (MacoNPV-A and MacoNPV-B) isolated from *M. configurata* belong to different species (Li et al., 2002a), which are sister groups (figure 25) within a large clade of Noctuidae-specific baculoviruses (figure 28). They have different biological activities in terms of virulence for *M. configurata* and alternative noctuid hosts (*Spodoptera exigua* and *Trichoplusia ni*). From a genomic insertion from XcGV the authors also predicted that MacoNPV-B might also be pathogenic for *Xestia c-nigrum* (Li et al., 2002a). My own phylogenetic study (figure NPVsp, section D) showed that viruses of the same species as MacoNPV-B were also isolated from *Mamestra brassicae* (MbNPV) and *Mythimna separata*. This combined with bioassay evidence (Li et al., 2002a) suggests that *Mamestra configurata NPV-B* has a host range including six species of Noctuid larvae. This is consistent with laboratory host range determination studies on MbNPV, which have shown its broad host range spanning 5 lepidopteran families (Doyle et al., 1990). My phylogenetic study also showed that the second virus species MacoNPV-A belongs to the same species as *Panolis flammea NPV* (figure 25). *Mamestra brassicae* is an alternative host for this virus (Hodgson et al., 2002). Although all this evidence is circumstantial and should be confirmed by bioassays, we have here two virus sister-species with different but overlapping host ranges. This might illustrate the results of a case of host range expansion, which could have been triggered by a horizontal gene transfer from XcGV to MacoNPV-B (Li et al., 2002a), leading to speciation. There is however no evidence that the other viruses of the MacoNPV-B species share the XcGV gene insertion.

Both *DNA polymerase* (Nielsen et al., 2002) and *polh* (figure 25) phylogenetic analyses have found that the baculovirus species *Malacosoma disstria* NPV and *Malacosoma californicum pluviale NPV* are closely related. These two viruses infect caterpillars from the same genus and belong to a clade of Lasiocampidae-specific viruses (figure 28). This might illustrate speciation after adaptation to a new host

species. Similar cases were found with baculoviruses infecting lymantriids of the genera *Orgyia* and *Lymantria*, where several closely related virus species are found to infect hosts from one genus (figure 28). Caterpillars of the species *Orgyia antiqua* have been found to be permissive to viruses isolated from three other species of the genus *Orgyia* (MNPV and SNPV of *O. pseudotsugata,* and NPVs of *O. mixta,* and *O. leucostigma*), as well as to its own virus (Richards et al., 1999). This cross-infectivity shows that even though these viruses belong to different species (figure 25) and might be adapted to their specific hosts, they have retained the capacity to infect hosts closely related to their own under laboratory conditions. Field studies would be necessary to determine whether the different *Orgyia* viruses are actually strictly specific in the wild to one host or if the species separation is rather due to geographic distribution, whereby each species reflects a particular region.

b) Ancestral Lepidopteran Hosts

The lepidopteran host mapping strongly suggested that the ancestral host of the lepidopteran baculoviruses belonged to the Noctuoidea. For all four topologies of *polh* and *lef-8* trees examined, all the MPR reconstructions of ancestral hosts agreed on the ancestral hosts superfamily to be Noctuoidea (figures 30-31). Furthermore within the Noctuoidea, the family level reconstructions showed that a species of the family Noctuidae was the most likely to have been the ancestral hosts of the ancestral lepidopteran baculovirus (figure 28, 31). Only one topology did not fully support this. For the polh-congruence tree the reconstruction was equivocal for the branch leading to the GVs, but had Noctuidae at the base of the NPV branch (not shown).

Despite the topological differences and low level of support at intermediate nodes (section D), both the *polh* and *lef-8* genes gave trees that agreed on both the superfamily and family of the baculovirus ancestral host. This shows that even if the lower branches of the phylogenies are poorly resolved there is still some structure within the trees. Although viruses infecting Noctuidae were the most represented (34% for the 94 taxa tree and 32% for the 68 taxa trees), they did not represent the majority of the viruses sampled in the phylogenies. Therefore, the results from the reconstruction of the ancestral host family cannot be entirely attributed to sample bias.

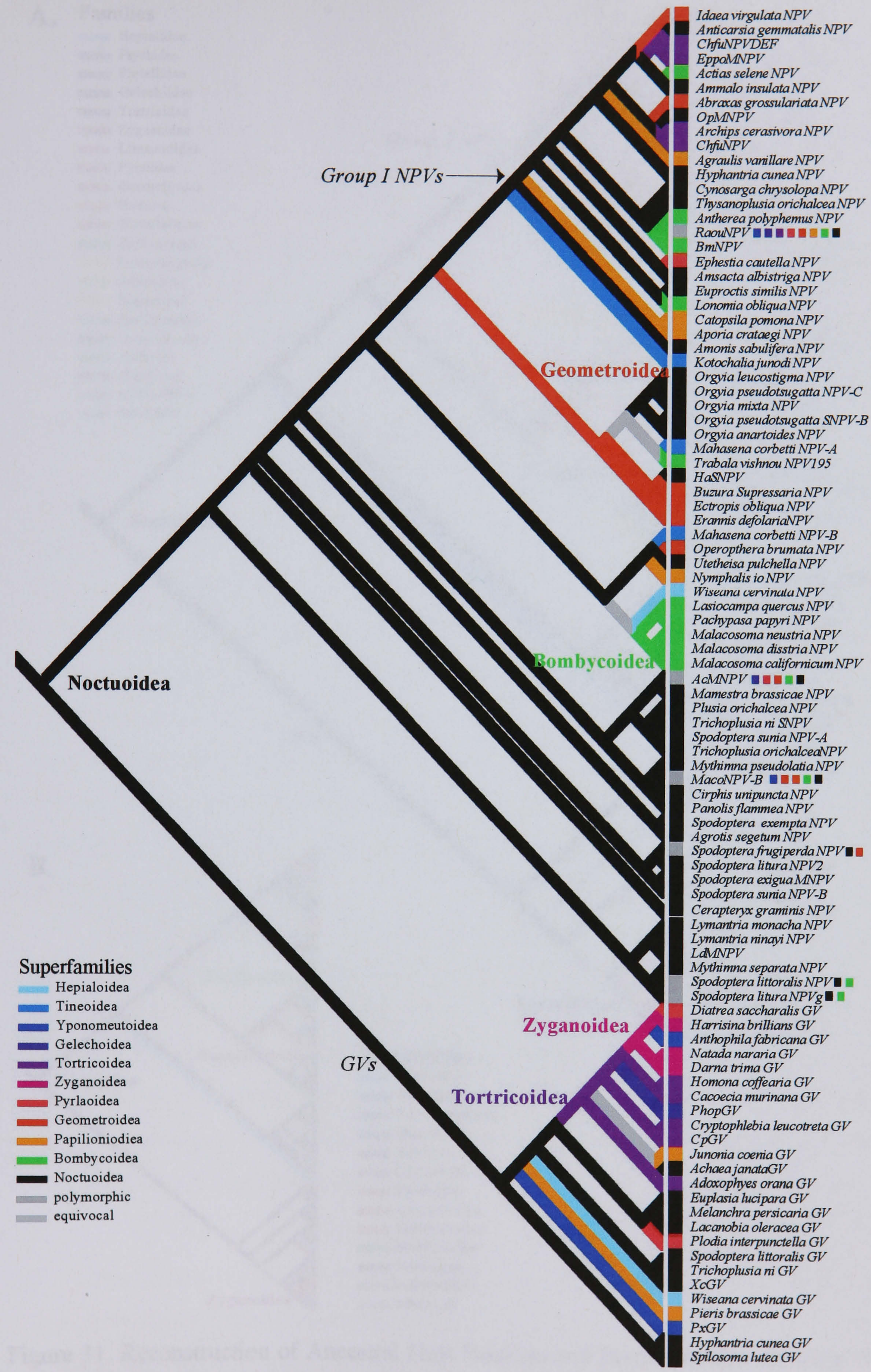


Figure 30: Reconstruction of Ancestral Host Super-Families on the *polyhedrin* Tree.

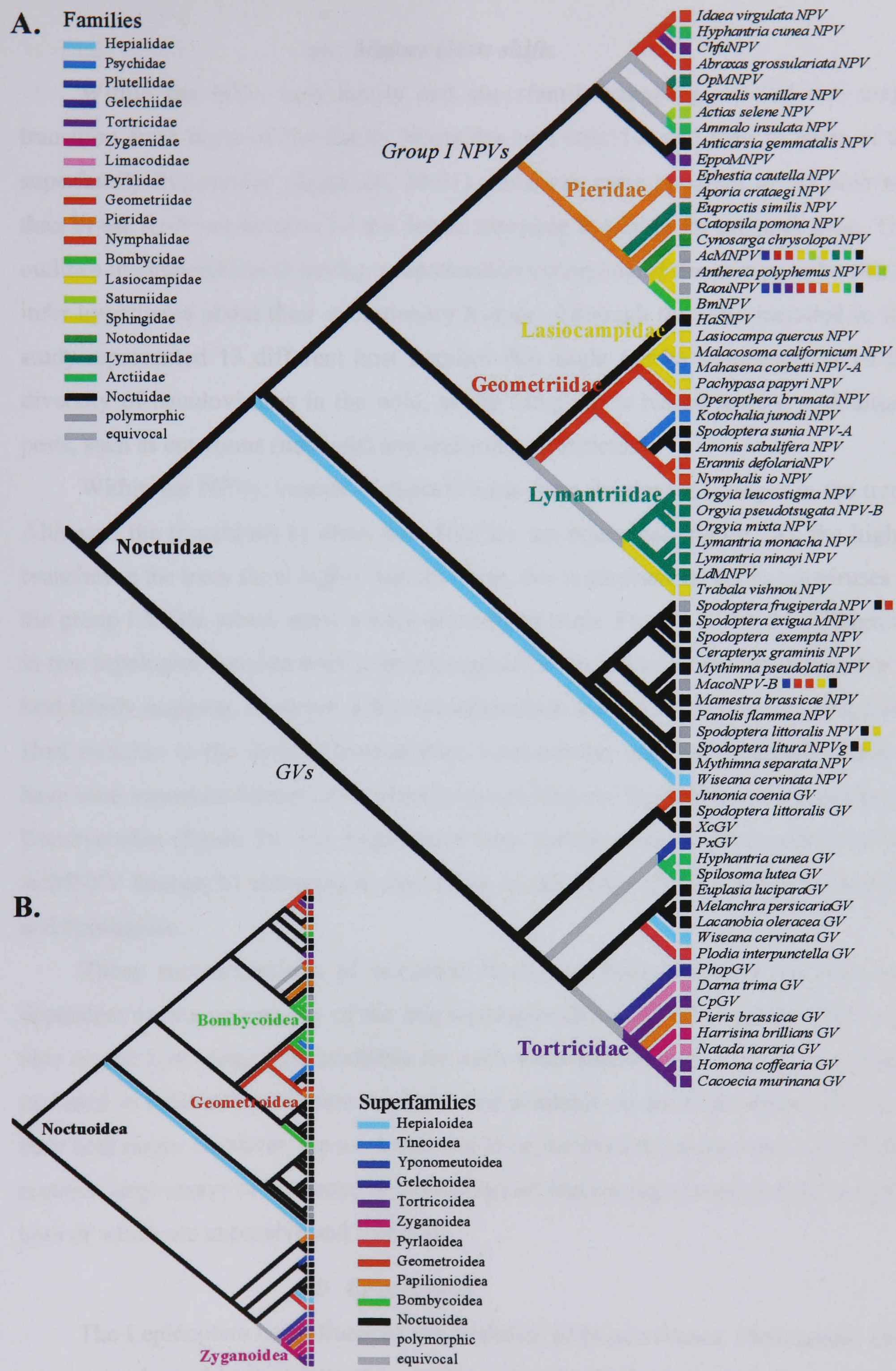


Figure 31: Reconstruction of Ancestral Host Families and Super-Families on the *lef-8* Tree. A. Mapping of the host families; B. Mapping of the host super-families.

c) Majors Hosts shifts

Within the GVs, host family and superfamily mapping showed one major transition from hosts of the family Noctuidae to Torticidae and then to hosts of the superfamily Zyganoidea (figure 28, 30-31). This was more apparent in the *polh* tree than in the *lef-8* tree because of the denser sampling available for the *polh* gene. This outlines the importance of having a representative sampling of the groups to be able to infer hypotheses about their evolutionary history. Although the GVs included in this study represented 13 different host families this might not be representative of the diversity of baculoviruses in the wild, as the sampling is biased toward agricultural pests, such as cutworms (noctuids) and leaf-rollers (tortricids).

Within the NPVs, viruses of noctuid hosts form the deeper lineages in the trees. Although the transitions to other host families are equivocal (figure 28), the higher branches in the trees show higher host diversity, this is emphasized for baculoviruses of the group I NPVs, which show a wide diversity of hosts. For the NPVs the differences in tree topologies between both gene phylogenies are responsible for the differences in host family mapping. However, a few reconstructions are conserved between both trees. Host switches to the family Geometriidae, Lymantriidae and Lasiocampidae seem to have been important factors of baculovirus diversification. Host shift to the superfamily Bombycoidea (figure 28, 31), might have been instrumental in the evolution of the AcMNPV lineage, by enlarging its host range to caterpillars of the families Saturniidae and Sphingidae.

These reconstructions of ancestral hosts and hosts-switches are not only dependent on the uncertainty of the tree topologies (Ronquist & Liljeblad, 2001), but also on the host range data available for each virus species. For most of the viruses included in these analyses there are little data available on their provenance let alone their host range. However, the work that would be involved to produce this kind of data requires large arrays of bioassays or collecting and sequencing numerous field samples, both of which are expensive and laborious.

d) Conclusions

The Lepidoptera has influenced the evolution of baculoviruses. Phylogenies from two genes showed host conservatism within the trees. Furthermore it is likely that the ancestral host of the lepidopteran baculoviruses belonged to the family Noctuidae.

To understand further the evolution of baculoviruses, it would be interesting to map the geographic distribution of virus species to see if biogeographic patterns

emerge. However, I suspect that such patterns might have been hidden by the worldwide spread of insect pests and of the viruses used as biological control agents. Another factor that could have influenced the evolution of baculoviruses are the hosts plants on which the host larvae feed. Baculoviruses might have had to adapt to persist on trees or shrubs and this could have triggered other levels of diversification.

Nevertheless, it is clear that ultimately baculoviruses have to be adapted to their hosts and that this is probably the factor that contributes the most to the patterns of evolutionary diversification.

F. Summary

The phylogenetic studies of baculovirus have provided the means to unravel their diversity. Sequences of the genes *ac22*, *lef-8* and *polyhedrin* were used to reconstruct phylogenetic trees. Some conflict was observed between the various gene trees, however most of it was attributed to low phylogenetic signal in poorly resolved parts of the trees. Overall the *Baculoviridae* can be divided into dipteran, hymenopteran and lepidopteran baculoviruses. Within the lepidopteran baculoviruses, further phylogenetic segregation was observed between the *Granulovirus* and *Nucleopolyhedrovirus* genera, with in addition a clear cluster of viruses forming the group I NPVs. A dense taxa sampling was achieved to encompass the diversity of the baculoviruses and their hosts. A number of isolates were found to cluster into baculovirus species units. This highlighted the diversity of baculovirus assemblages that can constitute distinct species.

The evolutionary history of baculoviruses has been characterised by host conservatism on several levels. Phylogenetic trees of baculoviruses, isolated from Lepidoptera, Diptera and Hymenoptera, showed clear divisions of the *Baculoviridae* according to the classification of their hosts at the level of insect orders. For the lepidopteran baculoviruses, phylogenetic analyses revealed a high degree of phylogenetic conservatism of host utilisation, whereby closely related viruses tend to infect more closely related hosts. Furthermore, reconstruction of ancestral host use suggested that the ancestral host of lepidopteran baculoviruses belonged to the family Noctuidae.

Sampling the phylogenetic diversity of the baculoviruses has been invaluable for unravelling the interrelationships of baculovirus field isolates. It has also highlighted the need for further studies to understand the evolution of baculoviruses in the field, including the interactions between strains and viral species.

IV. References

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

Ahrens, C. H., Russell, R. L. Q., Funk, C. J., Evans, J. T., Harwood, S. H. & Rohrmann, G. F. (1997). The sequence of the *Orgyia pseudotsugata* multicapsid nuclear polyhedrosis virus genome. *Virology* **229**, 381-399.

Albà, M. M., Das, R., Orengo, C. A. & Kellam, P. (2001). Genomewide Function Conservation and Phylogeny in Herpesviridae. *Genome Res.* **11**, 43-54.

Alm, R. A., Bina, J., Andrews, B. M., Doig, P., Hancock, R. E. W. & Trust, T. J. (2000). Comparative genomics of *Helicobacter pylori*: Analysis of the outer membrane protein families. *Infection and Immunity* **68**, 4155-4168.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990). Basic local alignment search tool. *J. Mol. Biol.* **215**, 403-410.

Arana, E. I., Albarino, C. G., O'Reilly, D., Ghiringhelli, P. D. & Romanowski, V. (2001). Generation of a recombinant *Anticarsia gemmatalis* multicapsid nucleopolyhedrovirus expressing a foreign gene under the control of a very late promoter. *Virus Genes* **22**, 363-372.

Argaud, O., Croizier, L., López-Ferber, M. & Croizier, F. (1998). Two key mutations in the host-range specificity domains of the *p143* gene of *Autographa californica* nucleopolyhedrovirus are required to kill *Bombyx mori* larvae. *J. Gen. Virol.* **79**, 931-935.

Arif, B., Lauzon, H. A. M., Pavlik, L., Lucarotti, C. J. & van Frankenhuyzen, K. (2002). Gene Organizationand genome sequencing of the readheaded pine sawfly, *Neodiprion lecontei*, nucleopolyhedrovirus. In *The world of microbes*, pp. 79. Edited by I. U. o. M. Societies. Paris: EDK Paris.

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

Balch, R. E. & Bird, F. T. (1944). A disease of the European spurce sawfly, *Gilpinia hercyniae* (htg), and its place in natural control. *Sci.Agric.* **25**, 65.

Barber, K. N., Kaupp, W. J. & Holmes, S. B. (1993). Specificity testing of the nuclear polyhedrosis-virus of the gypsy-moth, *Lymantria dispar*(L) (Lepidoptera, Lymantriidae). *Can. Entomol.* **125**, 1055-1066.

Barrett, J. W., Lauzon, H. A. M., Mercuri, P. S., Krell, P., Sohi, S. S. & Arif, B. (1996). The putative LEF-1 proteins from two distinct *Choristoneura fumiferana* multiple nucleopolyhedroviruses share domain homology to eukaryotic primases. *Virus Genes* **13**, 229.

Belyavskyi, M., Braunagel, S. & Summers, M. (1998). The structural protein ODV-EC27 of *Autographa californica* nucleopolyhedrovirus is a multifunctional viral cyclin. *Proc. Natl. Acad. Sci. USA* **95**, 11205-11210.

Benz, G. A. (1986). Introduction: Historical Perspective. In *The Biology of Baculoviruses*, pp. 1-35. Edited by R. R. Granados & B. A. Federici. Boca Raton, Florida: CRC Press.

Bergold, G. (1953). Insect viruses. *Adv. Virus Res.* **1**, 91-139.

Bideshi, D. K., Bigot, Y. & Federici, B. A. (2000). Molecular characterization and phylogenetic analysis of the *Harrisina brillians* granulovirus granulin gene. *Arch. Virol.* **145**, 1933-1945.

Black, B. C., Brennan, L. A., Dierks, P. M. & Gard, I. E. (1997). Commercialization of baculoviral insecticides. In *The Baculoviruses*, pp. 341-387. Edited by L. K. Miller. New York: Plenum Press.

Blanchette, M., Kunisawa, T. & Sankoff, D. (1999). Gene order breakpoint evidence in animal mitochondrial phylogeny. *J. Mol. Evol.* **49**, 193-203.

Blissard, G. W., Black, B., Crook, N., Keddie, B. A., Possee, R., Rohrmann, G. F., Theilmann, D. A. & Volkman, L. (2000). Family *Baculoviridae*. In *Virus Taxonomy: Seventh Report of the International Committee on Taxonomy of Viruses*, pp. 195-202. Edited by v. e. al. San Diego, USA: Academic Press.

Blissard, G. W., Quant, R. R. L., Rohrmann, G. F. & 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.

Blissard, G. W. & 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.

Blissard, G. W. & Rohrmann, G. F. (1990). Baculovirus diversity and molecular biology. *Ann. Rev. Entomol.* **35**, 127-55.

Blissard, G. W. & Wenz, J. R. (1992). Baculovirus gp64 envelope glycoprotein is sufficient to mediate pH-dependent membrane fusion. *J. Virol.* **66**, 6829-6835.

Boore, J. L. & Brown, W. M. (1998). Big trees from litle genomes: mitochondrial gene order as a phylogenetic tool. *Curr. Opin. Genet. Dev.* **8**, 668-674.

Boore, J. L., Collins, T. M., Stanton, D., Daehler, L. L. & Brown, W. M. (1995). Deducing the pattern of Arthropod phylogeny from mitochondrial DNA rearrangements. *Nature* **376**, 163-165.

Bork, P., Dandekar, T., Diaz-Lazcoz, Y., Eisenhaber, F., Huynen, M. & Yuan, Y. P. (1998). Predicting function: from genes to genomes and back. *J. Mol. Biol.* **283**, 707-725.

Bourque, G. & Pevzner, P. A. (2002). Genome-scale evolution: reconstructing gene orders in ancestral species. *Genome Res.* **12**, 26-36.

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

Braunagel, S. C., He, H., Ramamurthy, P. & 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.

Bulach, D. M., Kumar, C. A., Zaia, A., Liang, B. & Tribe, D. E. (1999). Group II nuclepolyhedrovirus subgroups revealed by phylogenetic analysis of polyhedrin and DNA polymerases gene sequences. *J. Invertebr. Pathol.* **73**, 59-73.

Burand, J. P. (1999). Nonoccluded baculovirses. In *Encyclopedia of Virology*, 2nd edn. Edited by R. G. Webster & A. Granoff. London: Academic Press.

Burden, J. P., Griffiths, C. M., Cory, J. S., Smith, P. & Sait, S. M. (2002). Vertical transmission of sublethal granulovirus infection in the indian meal moth, *Plodia interpunctella*. *Mol. Ecol.* **11**, 547-555.

Cameron, I. R. & Possee, R. D. (1989). Conservation of polyhedrin gene promoter function between *Autographa-californica* nuclear polyhedrosis virus and *Mamestra-brassicae* nuclear polyhedrosis virus. *Virus Res.* **12**, 183-200.

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

Chen, C. J. & Thiem, S. M. (1997). Differential infectivity of two *Autographa californica* nucleopolyhedrovirus mutants on three permissive cell lines is the result of lef-7 deletion. *Virology* **227**, 88-95.

Chen, X., Ijkel, W. F. J., Dominy, C., Zanotto, P., Hashimoto, Y., Faktor, O., Hayakawa, T., Wang, C.-H., Prekumar, A., Mathavan, S., Krell, P. J., Hu, Z. & Vlak, J. M. (1999). Identification, sequence analysis and phylogeny of the *lef-2* gene of *Helicoverpa armigera* single-nucleocapsid baculovirus. *Virus Res.* **65**, 21-32.

Chen, X., WF, I. J., Tarchini, R., Sun, X., Sandbrink, H., Wang, H., Peters, S., Zuidema, D., Lankhorst, R. K., Vlak, J. M. & Hu, Z. (2001). The sequence of the *Helicoverpa armigera* single nucleocapsid nucleopolyhedrovirus genome. *J. Gen. Virol.* **82**, 241-257.

Chen, X. W., Hu, Z. H., Jehle, J. A., Zhang, Y. Q. & Vlak, J. M. (1997). Analysis of the ecdysteroid UDP-glucosyltransferase gene of *Heliothis armigera* single-nucleocapsid baculovirus. *Virus Genes* **15**, 219-225.

Chen, X. W., Zhang, W. J., Wong, J., Chun, G., Lu, A., McCutchen, B. F., Presnail, J. K., Herrmann, R., Dolan, M., Tingey, S., Hu, Z. H. & Vlak, J. M. (2002). Comparative analysis of the complete genome sequences of *Helicoverpa zea* and *Helicoverpa armigera* single-nucleocapsid nucleopolyhedroviruses. *J. Gen. Virol.* **83**, 673-684.

Cheng, C. H., Liu, S. M., Chow, T. Y., Hsiao, Y. Y., Wang, D. P., Huang, J. J. & Chen, H. H. (2002). Analysis of the complete genome sequence of the Hz-1 virus suggests that it is related to members of the *Baculoviridae*. *J. Virol.* **76**, 9024-9034.

Cheng, X.-W., Garner, G. R. & Fescemeyer, H. W. (1998). Polyhedrin sequence determines the tetrahedral shape of occlusion bodies in *Thysanoplusia orichalcea* single-nucleocapsid nucleopolyhedrovirus. *J. Gen. Virol.* **79**, 2549-2556.

Cherry, R. (1993). Sericulture. *Bull. Entomol. Soc. Am.* **35**, 83-84.

Clarke, E. E., Tristem, M., Cory, J. S. & O'Reilly, D. R. (1996). Characterization of the ecdysteroid UDP-glucosyltransferase gene from *Mamestra brassicae* nucleopolyhedrosis virus. *J. Gen. Virol.* **77**, 2865-2871.

Clem, R. J. (1997). Regulation of programmed cell death by baculoviruses. In *The Baculoviruses*, pp. 237-261. Edited by L. K. Miller. New York: Plenum Press.

Clem, R. J. & Miller, L. K. (1994). Control of programmed cell death by the baculovirus genes *p35* and *iap*. *Mol. Cell. Biol.* **14**, 5212-5222.

Combes, C. (2001). Les associations du vivant - l'art d'être parasite. In *Nouvelle Bibliothèque Scientifique*, pp. 348. Paris: Flammarion.

Cory, J. S., Hails, R. S. & Sait, S. M. (1997). Baculovirus ecology. In *The Baculoviruses*, pp. 301-339. Edited by L. K. Miller. New York: Plenum Press.

Cory, J. S., Hirst, M. L., Sterling, P. A. & Speight, M. R. (2000). Narrow host range nucleopolyhedrovirus for control of the browntail moth (Lepidoptera : Lymantriidae). *Environ. Entom.* **29**, 661-667.

Cowan, P., Bulach, D., Goodge, K., Robertson, A. & Tribe, D. E. (1994). Nucleotide sequence of the polyhedrin gene region of *Helicoverpa zea* single nucleocapsid nuclear polyhedrosis virus: placement of the virus in lepidopteran nuclear polyhedrosis virus group II. *J. Gen. Virol.* **75**, 3211-3218.

Croizer, L., Taha, A., Croizer, G. & Lopez Ferber, M. (2002). Sequence analysis of the potato tuber moth, *Phthorimaea operculella*, granulovirus. In *XXXV Annual meeting of the SIP*, pp. 96. Edited by S. f. I. Pathology. Foz do Iguassu - Brazil.

Croizier, G. & Ribeiro, H. C. T. (1992). Recombination as a possible major cause of genetic heterogeneity in *Anticarsia gemmatalis* nuclear polyhedrosis virus populations. *Virus Res.* **26**, 183-196.

Cunningham, J. C. & Entwistle, P. F. (1981). Control of sawflies by baculovirus. In *MicrobialControl of Pests and plant diseases*. Edited by H. D. Burges. London: Academic Press.

Dandekar, T., Schuster, S., Snel, B., Huynen, M. & Bork, P. (1999). Pathway alignment: application to the comparative analysis of glycolytic enzymes. *Biochem. J.***343**, 115-124.

Dandekar, T., Snel, B., Huynen, M. & Bork, P. (1998). Conservation of gene order: a fingerprint of proteins that physically interact. *Trends Biochem. Sci.* **23**, 324-328.

Doyle, C. J., Hirst, M. L., Cory, J. S. & Entwistle, P. F. (1990). Risk assessment studies - detailed host range testing of wild-type cabbage moth, *Mamestra-brassicae* (Lepidoptera, Noctuidae), nuclear polyhedrosis virus. *Appl. Environ. Microbiol.* **56**, 2704-2710.

Drebot, M. A., Henchal, E., Hjelle, B., LeDuc, J. W., Repik, P. M., Roehrig, J. T., Schamljohn, C. S., Shope, R. E., Tesh, R. B., Weaver, S. C. & Calisher, C. H.

(2002). Improved clarity of meaning from the use of both formal species names and common (vernacular) virus names in virological literature. *Arch. Virol.* **147**, 2465-2471.

Ehrlich, P. R. & Raven, P. H. (1964). Butterflies and plants: a study in coevolution. *Evolution* **18**, 586-608.

Eisen, J. A. (2002). Brouhaha over the other yeast. *Nature* **415**, 845-848.

Eisenberg, D., Marcotte, E. M., Xenarios, I. & Yeates, T. O. (2000). Protein function in the post-genomic era. *Nature* **405**, 823-826.

Eldridge, R., Li, Y. & Miller, L. K. (1992). Characterization of a baculovirus gene encoding a small conotoxin-like polypeptide. *J. Virol.* **66**, 6563-6571.

Evans, J. T., Leisy, D. J. & Rohrmann, G. F. (1997). Characterization of the interaction between the baculovirus replication factors LEF-1 and LEF-2. *J. Virol.* **71**, 3114-3119.

Faulkner, P., Kuzio, J., Williams, G. V. & Wilson, J. A. (1997). Analysis of p74, a PDV envelope protein of *Autographa californica* nucleopolyhedrovirus required for occlusion body infectivity *in vivo*. *J. Gen. Virol.* **78**, 3091-3100.

Federici, B. A. (1997). Baculovirus pathogenesis. In *The Baculoviruses*, pp. 33-56. Edited by L. K. Miller. New York and London: Plenum Press.

Federici, B. A. & Stern, V. M. (1990). Replication and occlusion of a granulosis virus in larval and adult midgut epithelium of the western grapeleaf skeletonizer, *Harrisina brillians*. *J. Invertebr. Pathol.* **56**, 401-414.

Felsenstein, J. (1978). Cases in which parsimony or compatibility methods will be positively misleading. *Syst. Biol.* **27**, 401-410.

Felsenstein, J. (2000.). PHYLIP. Phylogeny Inference Package., 3.6 edn. Washington.

Fitz-Gibbon, S. T. & House, C. H. (1999). Whole genome-based phylogenetic analysis of free living microorganisms. *Nucleic Acids Res.* **27**, 4218-4222.

Fraser, M. J., Cary, L., Boonvisudhi, K. & Wang, H.-G. H. (1995). Assay for movement of lepidopteran transposon IFP2 in insect cells using a baculovirus genome as a target DNA. *Virology* **211**, 397-407.

Friesen, P. D. (1997). Regulation of baculovirus early gene expression. In *The Baculoviruses*, pp. 141-170. Edited by L. K. Miller. New York: Plenum Press.

Friesen, P. D. & Nissen, M. S. (1990). Gene organization and transcription of TED, a lepidopteran retrotransposon integrated within the baculovirus genome. *Mol. Cell. Biol.* **10**, 3067-3077.

Fuchs, L. Y., Woods, M. S. & Weaver, R. F. (1983). Viral transcription during *Autographa californica* nuclear polyhedrosis virus infection: a novel RNA polymerase induced in infected *Spodoptera frugiperda* cells. *J. Virol.* **43**, 641-646.

Funk, C. J., Braunagel, S. C. & Rohrman, G. F. (1997). Baculovirus structure. In *The Baculoviruses*, pp. 7-27. Edited by L. K. Miller. New York and London: Plenum Press.

Futuyama, D. J. (1998). Evolutionary Biology, 3rd edn. Sunderland, Massachusetts: Sinauer Associates Inc.

Fuxa, J. R., Richter, A. R., Ameen, A. O. & Hammock, B. D. (2002). Vertical Transmission of TnSNPV, TnCPV, AcMNPV, and possibly recombinant NPV in *Trichosplusia ni*. *J. Invertebr. Pathol.* **79**, 44-50.

Galperin, M. Y. & Koonin, E. V. (2000). Who's your neighbor? New computational approaches for functional genomics. *Nature Biotech.* **18**, 609-613.

Gascuel, O. (1997). BIONJ: an improved version of the NJ algorithm based on a simple model of sequence data. *Mol. Biol. Evol.* **14**, 685-695.

Gomi, S., Majima, K. & Maeda, S. (1999). Sequence analysis of the genome of *Bombyx mori* nucleopolyhedrovirus. *J. Gen. Virol.* **80**, 1323-1337.

Guarino, L., Jin, J. & Dong, W. (1998a). Guanylyltransferase activity of the LEF-4 subunit of baculovirus RNA polymerase. *J. Virol* **72**, 10003-10010.

Guarino, L. A., Xu, B., Jin, J. & Dong, W. (1998b). A virus-encoded RNA polymerase purified from baculovirus-infected cells. *J. Virol.* **72**, 7985-7991.

Hang, X., Dong, W. & Guarino, L. A. (1995). The *lef-3* gene of *Autographa californica* nuclear polyhedrosis virus encodes a single-stranded DNA-binding protein. *J. Virol.* **69**, 3924-3928.

Hannenhalli, S., Chappey, C., Koonin, E. V. & Pevzner, P. A. (1995). Genome sequence comparison and scenarios for gene rearrangements - a test-case. *Genomics* **30**, 299-311.

Harrison, R. L. & Boning, B. C. (2000). Use of scorpion neurotoxins to improve the insecticidal activity of *Rachiplusia ou* multicapsid nucleopolyhedrovirus. *Biol. Contr.* **17**, 191-201.

Harrison, R. L. & Bonning, B. (2002). The sequence of *Rachiplusia ou* multi-nucleocapsid nucleopolyhedrovirus genome. In *XXXV Annual meeting of the SIP*, pp. 96. Edited by S. f. I. Pathology. Foz do Iguassu - Brazil.

Hasegawa, M., Kishino, K. & Yano, T. (1985). Dating the human-ape splitting by a molecular clock of mitochondrial DNA. *J. Mol. Evol.* **22**, 160-174.

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

Hayakawa, T., Ko, R., Okano, K., Seong, S. I., Goto, C. & Maeda, S. (1999). Sequence analysis of the *Xestia c-nigrum* granulovirus genome. *Virology* **262**, 277-97.

Hayakawa, T., Rohrmann, G. F. & Hashimoto, Y. (2000). Patterns of genome organization and content in lepidopteran baculoviruses. *Virology* **278**, 1-12.

Herniou, E., Martin, J., Miller, K., Cook, J., Wilkinson, M. & Tristem, M. (1998). Retroviral diversity and distribution in vertebrates. *J. Virol.* **72**, 5955-5966.

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

Herniou, E. A., Olszewski, J., Cory, J. & O'Reilly, D. (2003). The genome sequence and evolution of baculoviruses. *Ann. Rev. Entomol.* **48**, 211-234.

Hodgson, D. J., Vanbergen, A. J., Hartley, S. E., Hails, R. S. & Cory, J. S. (2002). Differential selection of baculovirus genotypes mediated by different species of host food plant. *Ecology Letters* **5**, 512-518.

Hodgson, D. J., Vanbergen, A. J., Watt, A. D., Hails, R. S. & Cory, J. S. (2001). Phenotypic variation between naturally co-existing genotypes of a Lepidopteran baculovirus. *Evol. Ecol. Res.* **3**, 687-701.

Holmes, E. C. (1999). Molecular phylogenies and the genetic structure of viral populations. In *Evolution in Health & Disease*, pp. 173-182. Edited by S. C. Stearns. New York: Oxford University Press.

Hostetter, D. L. & Puttler, B. (1991). A new broad host spectrum nuclear polyhedrosis virus isolated from a celery looper, *Anagrapha falcifera* (Kirby), (Lepidoptera: Noctuidae). *Environ. Entom.* **20**, 1480-1488.

Hu, Z. H., Arif, B. M., Jin, F., Martens, J. W. M., Chen, X. W., Sun, J. S., Zuidema, D., Goldbach, R. W. & Vlak, J. M. (1998). Distinct gene arrangement in the *Buzura suppressaria* single-nucleocapsid nucleopolyhedrovirus genome. *J. Gen. Virol.* **79**, 2841-2851.

Huelsenbeck, J. P. (1991). Tree-length distribution skewness: An indicator of phylogenetic information. *Syst. Zool.* **40**, 257–270.

Hyink, O., Dellow, R. A., Olsen, M. J., Caradoc-Davies, K. M. B., Drake, K., Herniou, E. A., Cory, J. S., O'Reilly, D. R. & Ward, V. K. (2002). Whole genome analysis of the *Epiphyas postvittana* nucleopolyhedrovirus. *J. Gen. Virol.* **83**, 959-973.

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

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

Jehle, J. A. (2002). The sequence of *Cryptophlebia leucotreta* granulovirus. In *XXXV Annual meeting of the SIP*, pp. 96. Edited by S. f. I. Pathology. Foz do Iguassu - Brazil.

Jehle, J. A., Fritsch, E., Nickel, A., Huber, J. & Backhaus, H. (1995). TC14.7: a novel lepidopteran transposon found in *Cydia pomonella* granulosis virus. *Virology* **207**, 369-379.

Jehle, J. A., Nickel, A., Vlak, J. M. & Backhaus, H. (1998). Horizontal escape of the novel Tc10-like lepidopteran transposon TCp3.2 into *Cydia pomonella* granulovirus. *J. Mol. Evol.* **46**, 215-224.

Jin, J., Dong, W. & Guarino, L. A. (1998). The LEF-4 subunit of baculovirus RNA polymerase has RNA 5'- triphosphatase and ATPase activities. *J. Virol.* **72**, 10011-9.

Jin, T., Qi, Y., Liu, D. & Su, F. (1999). Nucleotide sequence of a 5892 base pairs fragment of the LsMNPV genome and phylogenetic analysis of LsMNPV. *Virus Genes* **18**, 265-276.

Kang, W., Tristem, M., Maeda, S., Crook, N. E. & O'Reilly, D. R. (1998). Identification and characterization of the *Cydia pomonella* granulovirus cathepsin and chitinase genes. *J. Gen. Virol.* **79**, 2283-2292.

Kikhno, I. M., Gutierrez, S., Croizer, L., Croizer, G. & Lopez Ferber, M. Characterisation of *pif*, a gene required for the *per os* infectivity of Spodoptera littoralis nucleopolyhedrovirus. *j. Gen. Virol.* **83,** 3013-3022.

Kishino, H. & Hasegawa, M. (1989). Evaluation of maximum likelihood estimate of the evolutionary tree topologies from DNA sequence data, and the branching order in Hominoidea. *J. Mol. Evol.* **29**, 170-179.

Kitching, I. J. & Rowlins, J. E. (1999). Noctuoidea. In *Handbook of Zoology. Lepidoptera, volume 1: Systematic and Evolution*, pp. 355-401. Edited by N. P. Kristensen. Berlin: W. de Gruyter.

Ko, R., Okano, K. & Maeda, S. (2000). Structural and functional analysis of the *Xestia c-nigrum* granulovirus matrix metalloproteinase. *J. Virol.* **74**, 11240-11246.

Kondo, A. & Maeda, S. (1991). Host range expansion by recombination of the baculoviruses *Bombyx mori* nuclear polyhedrosis virus and *Autographa californica* nuclear polyehdrosis virus. *J. Virol.* **65**, 3625-3632.

Kool, M., Ahrens, C. H., Goldbach, R. W., Rohrmann, G. F. & Vlak, J. M. (1994). Identification of genes involved in DNA replication of *Autographa californica* baculovirus. *Proc. Natl. Acad. Sci. U.S.A.* **91**, 11212-11216.

Koonin, E. V. (2000). How many genes can make a cell: the minimal-gene-set concept. *Ann. Rev. Genom. Hum. Genet.* **1**, 99-116.

Koonin, E. V., Aravind, L. & Kondrashov, A. S. (2000). The impact of comparative genomics on our understanding of evolution. *Cell* **101**, 573-576.

Koufopanou, V., Burt, A. & Taylor, J. W. (1997). Concordance in gene genealogies reveals reproductive isolation in the pathogeni fungus *Coccidioides immitis*. *Proc. Natl. Acad. Sci U.S.A.* **94**, 5478-5482.

Kreitman, M. & Gaasterland, T. (2001). Genomes and evolution - ships in the night? *Curr. Opin. Genet. Dev.* **11**, 609-6011.

Kukan, B. (1999). Vertical transmission of nucleopolyherdovirus in insects. *J. Invertebr. Pathol.* **74**, 103-111.

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

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

Lacey, L. A., Vail, P. V. & Hoffmann, D. F. (2002). Comparative activity of baculoviruses against the codling moth *Cydia pomonella* and three other tortricid pests of tree fruit. *J. Invertebr. Pathol.* **80**, 64-68.

Le, T. H., Blair, D., Agatsuma, T., Humair, P. F., Campbell, N. J. H., Iwagami, M., Littlewood, D. T. J., Peacock, B., Johnston, D. A., Bartley, J., Rollinson, D., Herniou, E. A., Zarlenga, D. S. & McManus, D. P. (2000). Phylogenies inferred

from mitochondrial gene orders - A cautionary tale from the parasitic flatworms. *Mol. Biol. Evol.* **17**, 1123-1125.

Lee, H. H. & Miller, L. K. (1978). Isolation of genotypic variants of *Autographa californica* nuclear polyhedrosis virus. *J. Virol.* **27**, 754-767.

Lewin, B. (2000). Genes VII, pp. 990. New York: Oxford University Press.

Li, C. B., Li, Z. F., Yan, Q. S., Hu, G. D. & Pang, Y. (2001). Cloning and sequencing of *Spodoptera litura* multicapsid nucleopolyhedrovirus gp37 gene cluster. *Prog. Biochem. Biophys.***28**, 677-682.

Li, L. L., Donly, C., Li, Q. J., Willis, L. G., Keddie, B. A., Erlandson, M. A. & Theilmann, D. A. (2002a). Identification and genomic analysis of a second species of nucleopolyhedrovirus isolated from *Mamestra configurata*. *Virology* **297**, 226-244.

Li, Q. J., Donly, C., Li, L. L., Willis, L. G., Theilmann, D. A. & Erlandson, M. (2002b). Sequence and organization of the *Mamestra configurata* nucleopolyhedrovirus genome. *Virology* **294**, 106-121.

Lin, J. & Gerstein, M. (2000). Whole-genome trees based on the occurrence of folds and orthologs: Implications for comparing genomes on different levels. *Genome Res.* **10**, 808-818.

Liu, G. & Carstens, E. B. (1999). Site-directed mutagenesis of the AcMNPV p143 gene: effects on baculovirus DNA replication. *Virology* **253**, 125-36.

Liu, J. & Maruniak, J. E. (1999). Molecular characterization of genes in the GP41 region of baculoviruses and phylogenetic analysis based upon GP41 and polyhedrin genes. *Virus Res* **64**, 187-196.

Lopez-Ferber, M., Argaud, O., Croizier, L. & Croizier, G. (2001). Diversity, distribution and mobility of *bro* gene sequences in *Bombyx mori* nucleopolyhedrovirus. *Virus genes* **22**, 247-254.

Lopez-Vaamonde, C., Godfray, H. C. J. & Cook, J. M. (2003). Evolutionary dynamics of host plant utilisation in a genus of leaf-mining moths. *Evolution* ***in press***.

Lu, A. & 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. *J. Virol.* **68**, 6710-6718.

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

Luque, T., Finch, R., Crook, N., O'Reilly, D. R. & Winstanley, D. (2001). The complete sequence of the *Cydia pomonella* granulovirus genome. *J. Gen. Virol.* **82**, 2531-2547.

Maddison, D. R. & Maddison, W. R. (2000). MacClade 4. Sunderland, Massachusetts: Sinauer Associates.

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

Malik, H. S., Henikoff, S. & Eickbush, T. H. (2000). Poised for contagion: evolutionary origins of the infectious abilities of invertebrate retroviruses. *Genome Res.* **10**, 1307-1318.

Marcotte, E. M., Pellegrini, M., Ng, H. L., Rice, D. W., Yeates, T. O. & Eisenberg, D. (1999). Detecting protein function and protein-protein interactions from genome sequences. *Science* **285**, 751-753.

Martens, J. W., Knoester, M., Weijts, F., Groffen, S. J., Hu, Z., Bosch, D. & Vlak, J. M. (1995). Characterization of baculovirus insecticides expressing tailored *Bacillus thuringiensis* CryIA(b) crystal proteins. *J. Invertebr. Pathol.* **66**, 249-57.

Martignoni, M. E. & Iwai, P. J. (1986). A catalogue of viral diseases of insects, mites, and ticks. 4th ed. General Technical Report PNW-195.Portland, OR: U.S. Department of Agriculture, Forest Service, Pacific Northwest Research Station; 51p.

Matthews, H. J., Smith, I. & Edwards, J. P. (2002). Lethal and sublethal effects of a granulovirus on the tomato moth *Lacanobia oleracea. J. Invertebr. Pathol.* **80,** 73-80.

Mayr, E. (1982). The growth of biological thought. Diversity, evolution, and inheritance. Edited by E. Mayr. Cambridge, MA: Harvard University Press.

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

McDougal, V. V. & Guarino, L. A. (2000). The *Autographa californica* nuclear polyhedrosis virus p143 gene encodes a DNA helicase. *J. Virol.* **74**, 5273-5279.

Mcgeoch, D. J., Dolan, A. & Ralph, A. C. (2000). Toward a comprehensive phylogeny for mammalian and avian herpesviruses. *J. Virol.* **74**, 10401-10406.

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

Mikhailov, V. S. & Rohrmann, G. F. (2002). Baculovirus replication factor LEF-1 is a DNA primase. *J. Virol.* **76**, 2287-2297.

Miller, L. K. (1997). The Baculoviruses. In *The Viruses*, pp. 447. Edited by H. Fraenkel-Conrat & R. R. Wagner. New York: Plenum Press.

Mitchell, A., Mitter, C. & Regier, J. C. (2000). More taxa or more characters revisited: Combining data from nuclear protein-encoding genes for phylogenetic analyses of Noctuoidea (Insecta : Lepidoptera). *Syst. Biol.* **49**, 202-224.

Montague, M. G. & Hutchison, C. A. (2000). Gene content phylogeny of herpesviruses. *Proc. Natl. Acad. Sci. U. S. A.* **97**, 5334-5339.

Morse, M. A., Marriott, A. C. & Nuttall, P. A. (1992). The glycoprotein of Thogoto virus (a tick-borne orthomyxo-like virus) is related to the baculovirus glycoprotein gp64. *Virology* **186**, 640-646.

Moscardi, F. (1999). Assessment of the application of baculoviruses for control of lepidoptera. *Annu. Rev. Entomol.* **44**, 257-289.

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

Munoz, D., Vlak, J. M. & Caballero, P. (1997). In vivo recombination between two strains of the genus *Nucleopolyhedrovirus* in its natural host, *Spodoptera exigua*. *Appl. Environ. Microb.* **63**, 3025-3031.

Myers, J. H., Malakar, R. & Cory, J. S. (2000). Sublethal nucleopolyhedrovirus infection effects on female pupal weight, egg mass size, and vertical transmission in gypsy moth (Lepidoptera : Lymantriidae). *Environ. Entomol.* **29**, 1268-1272.

Nakai, M., Goto, C., Kang, W. K., Shikata, M., Ishii, T., Takatsuka, J., Okuno, S. & Kunimi, Y. (2002). Characteristics and genome sequence of *Adoxophyes honmai* nucleopolyhedrovirus. In *XXXV Annual meeting of the SIP*, pp. 96. Edited by S. f. I. Pathology. Foz do Iguassu - Brazil.

Nielsen, C. B., Cooper, D., Short, S. M., Myers, J. H. & Suttle, C. A. (2002). DNA polymerase gene sequences indicate western and forest tent caterpillar viruses

form a new taxonomic group within baculoviruses. *J. Invertebr. Pathol.* **81**, 131-147.

Nikaido, M., Kawai, K., Cao, Y., Harada, M., Tomita, S., Okada, N. & Hasegawa, M. (2001). Maximum likelihood analysis of the complete mitochondrial genomes of eutherians and a reevaluation of the phylogeny of bats and insectivores. *J. Mol. Evol.* **53**, 508–516.

O'Reilly, D. R. (1997). Auxiliary genes of baculoviruses. In *The Baculoviruses*, pp. 267-300. Edited by L. K. Miller. New York: Plenum Press.

O'Reilly, D. R., Miller, L. K. & Luckow, V. A. (1992). Baculovirus expression vectors - a laboratory manual. New York: Oxford University Press.

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

Olszewski, J. & Miller, L. K. (1997b). A role for baculovirus GP41 in budded virus production. *Virology* **233**, 292-301.

Page, R. D. M. (1995). TreeMap. Glasgow.

Page, R. D. M. (1996). TreeView: An application to display phylogenetic trees on personal computers. *Comput. Appl. Biosci.* **12**, 357-358.

Paillot, A. (1927). Etude sur la grasserie (polyhédrie). In *2° congresso Serico Europeo*, pp. 10. Milano.

Pang, Y., Yu, J. X., Wang, L. H., Hu, X. H., Bao, W. D., Li, G., Chen, C., Han, H., Hu, S. N. & Yang, H. M. (2001). Sequence analysis of the *Spodoptera litura* multicapsid nucleopolyhedrovirus genome. *Virology* **287**, 391-404.

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

Pasteur, L. (1870). Etudes sur la maladie des vers a soie. In *Oeuvres de Pasteur*, 1926 edn. Edited by P. Vallery-Radot. Paris: Masson.

Payne, C. C. (1986). Insect pathogenic viruses as pest control agents. In *Biological Plant and Health Protection*, pp. 183-200. Edited by J. M. Franz. Stuttgart; New York: G. Fischer Verlag.

Pearson, M. N., Groten, C. & Rohrmann, G. F. (2000). Identification of the *Lymantria dispar* nucleopolyhedrovirus envelope fusion protein provides evidence for a phylogenetic division of the Baculoviridae. *J. Virol.* **74**, 6126-6131.

Pearson, M. N. & Rohrmann, G. F. (1998). Characterisation of a baculovirus-encoded ATP-dependent DNA ligase. *J. Virol.* **72**, 9142-9149.

Pearson, M. N. & Rohrmann, G. F. (2002). Transfer, incorporation, and substitution of envelope fusion proteins among members of the *Baculoviridae*, *Orthomyxoviridae*, and *Metaviridae* (insect retrovirus) Families. *J. Virol* **76**, 5301-5304.

Pearson, M. N., Russell, R. L. Q. & Rohrmann, G. F. (2002). Transcriptional mapping of two genes encoding baculovirus envelope-associated proteins. *J. Gen. Virol.* **83**, 937-943.

Pearson, M. N., Russell, R. L. Q. & Rohrmann, G. F. (2001). Characterization of a baculovirus-encoded protein that is associated with infected-cell membranes and budded virions. *Virology* **291**, 22-31.

Pearson, M. N., Russell, R. L. Q., Rohrmann, G. F. & Beaudreau, G. S. (1988). p39, A major baculovirus structural protein: Immunocytochemical characterization and genetic location. *Virology* **167**, 407-413.

Pellegrini, M., Marcotte, E. M., Thompson, M. J., Eisenberg, D. & Yeates, T. O. (1999). Assigning protein functions by comparative genome analysis: Protein phylogenetic profiles. *Proc. Natl. Acad. Sci. USA* **96**, 4285-4288.

Posada, D. & Crandall, K. A. (1998). MODELTEST: testing the model of DNA substitution. *Bioinformatics* **14**, 817-818.

Posada, D. & Crandall, K. A. (2001). Selecting the best-fit model of nucleotide substitution. *Syst. Biol.* **50**, 580-601.

Price, P. W. (2002). Species interactions and the evolution of biodiversity. In *Plant-Animal Interactions an Evolutionary Approach*, pp. 3-25. Edited by C. M. Herrera & O. Pellmyr. Oxford: Blackwell Publishing.

Ramasamy, P., Rajan, P. R., Purushothaman, V. & Brennan, G. P. (2000). Ultrastructure and pathogenesis of *Monodon baculovirus* (PmSNPV) in cultured larvae and natural brooders of *Peneaus monodon*. *Aquaculture* **184**, 45-66.

Ramoska, W. A., Stairs, G. R. & Hink, W. F. (1975). UV light activation of insect nuclear polyhedrosis virus. *Nature* **253**, 628-629.

Reichard, P. (2002). Ribonucleotide reductases: The Evolution of allosteric regulation. *Arch. Biochem. Biophys.* **397**, 149-155.

Reynaud (1824). Des vers à soie, et de leur éducation selon la pratique des Cévennes; suivi d'un précis sur les divers produits de la soie, et sur la manière de tirer les

fantaisies et les filoselles; avec des notions sur la fabrique des bas de Ganges, pp. 372. Paris: Libr. du Commerce.

Richards, A., Speight, M. R. & Cory, J. S. (1999). Characterization of a nucleopolyhedrovirus from the vapourer moth, *Orgyia antiqua* (Lepidoptera Lymantriidae). *J. Invertebr. Pathol.* **74**, 137-142.

Rodriguez, F., Oliver, J. F., Marin, A. & Medina, J. R. (1990). The general stochastic model of nucleotide substitution. *J. Theor. Biol.* **142**, 485-501.

Rohrmann, G. F. (1999). Nuclear Polyhedrosis Viruses. In *Encyclopedia of Virology*, 2nd edn. Edited by R. G. Webster & A. Granoff. London: Academic Press.

Rohrmann, G. F. & Karplus, P. A. (2001). Relatedness of baculovirus and gypsy retrotransposon envelope proteins. *BMC Evolutionary Biology* **1**.

Rohrmann, G. F., Pearson, M. N., Bailey, T. J., Becker, R. R. & Beaudreau, G. S. (1981). N-terminal polyhedrin sequences and occluded baculovirus evolution. *J. Mol. Evol.* **17**, 329-333.

Roncarati, R. & Knebel-Mörsdorf, D. (1997). Identification of the early actin-rearrangement-inducing factor gene, *arif-1*, from *Autographa californica* multicapsid nuclear polyhedrosis virus. *J. Virol.* **71**, 7933-7941.

Ronquist, F. & Liljeblad, J. (2001). Evolution of the gall wasp-host plant association. *Evolution* **12**, 2503-2522.

Rothman, L. D. & Myers, J. H. (1996). Debilitating effects of viral diseases on host lepidoptera. *J. Invertebr. Pathol.* **67**, 1-10.

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

Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989). Molecular cloning: a laboratory manual, second edn. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.

Sanderson, M. J. & Shaffer, H. B. (2002). Troubleshooting molecular phylogenetic analyses. *Annu. Rev. Ecol. Syst.* **33**, 49-72.

Scoble, M. J. (1992). The Lepidoptera: form, function and diversity. In *Natural History Museum Publications*, pp. 404. New York: Oxford University Press.

Shimodaira, H. & Hasegawa, M. (1999). Multiple comparisons of log-likelihoods with applications to phylogenetic inference. *Mol. Biol. Evol.* **16**, 1114-1116.

Sicheritz-Pontén, T. & Andersson, S. G. E. (2001). A phylogenomic approach to microbial evolution. *Nucleic Acids Res.* **29**, 545-552.

Snel, B., Bork, P. & Huynen, M. A. (1999). Genome phylogeny based on gene content. *Nature Genet.* **21**, 108-110.

Snel, B., Bork, P. & Huynen, M. A. (2002). Genomes in Flux: the Evolution of Archeal and Proteobacterial Gene Content. *Genome Res.* **12**, 17-25.

Stairs, G. (1971). Use of viruses for microbial control of insects. In *Microbial Control of Insects and Mites*, pp. 97-124. Edited by H. Barges & N. Hussey. New York:London: Academic Press.

Stiles, B. & Himmerich, S. (1998). *Autographa californica* NPV isolates: restriction endonuclease analysis and comparative biological activity. *J. Invertebr. Pathol.* **72**, 174-177.

Swofford, D. L. (2001.). PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). 4. edn. Sunderland, Massachusetts.: Sinauer Associates.

Takahashi, K. & Nei, M. (2000). Efficiencies of fast algorithms of phylogenetic inference under the criteria of maximum parsimony, minimum evolution, and maximum likelihood when a large number of sequences are used. *Mol. Biol. Evol.* **17**, 1251-1258.

Tamames, J., Gonzalez-Moreno, M., Mingorance, J., Valencia, A. & Vicente, M. (2001). Bringing gene order into bacterial shape. *Trends Genet.* **17**, 124-126.

Tekaia, F., Lazcano, A. & Dujon, B. (1999). The genomic tree as revealed from whole proteome comparisons. *Genome Res.* **9**, 550-557.

Telford, M. J. (2000). Evidence for the derivation of the drosophila *fushi tarazu* gene from a Hox gene orthologous to lophotrochozoan Lox5. *Curr. Biol.* **10**, 349–352.

Terra, W. R. & Ferreira, C. (1994). Insect digestive enzymes - properties, compartimentalization and function. *Comp. Biochem. Physiol. B Biochem. & Mol. Biol.* **109**, 1-62.

Thiem, S. M., Du, X. L., Quentin, M. E. & Berner, M. M. (1996). Identification of a baculovirus gene that promotes *Autographa californica* nuclear polyhedrosis virus replication in a nonpermissive insect cell line. *J. Virol.* **70**, 2221-2229.

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

Thompson, J. D., Higgins, D. G. & 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 Res.* **22**, 4673-4680.

Thompson, J. N. (2002). Plant-animal interactions: future directions. In *Plant-Animal Interactions an Evolutionary Approach*, pp. 236-247. Edited by C. M. Herrera & O. Pellmyr. Oxford: Blackwell Publishing.

Thornton, J. W. & DeSalle, R. (2000). Gene family evolution and homology: Genomics meets phylogenetics. *Ann. Rev. Genom. Hum. Genet.* **1**, 41-+.

Tillier, E. R. M. & Collins, R. A. (2000). Genome rearrangement by replication-directed translocation. *Nature Genet.* **26**, 195-197.

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

Trudeau, D., Washburn, J. O. & Volkman, L. E. (2001). Central role of hemocytes in *Autographa californica* M nucleopolyhedrovirus pathogenesis in *Heliothis virescens* and *Helicoverpa zea*. *J. Virol.* **75**, 996-1003.

Tweeten, K. A., Bulla, L. A., Jr. & Consigli, R. A. (1980). Characterization of an extremely basic protein derived from granulosis virus nucleocapsids. *J. Virol.* **33**, 866-876.

Vail, P. V., Hostetter, D. L. & Hoffmann, D. F. (1999). Development of the multi-nucleocapsid nucleopolyhderoviruses (MNPVs) infectious to loopers (Lepidoptera: Noctuidae: Plusiinae) as microbial control agents. *Int. Pest. Manage. Rev.* **4**, 231-257.

van Hulten, M. C., Witteveldt, J., Peters, S., Kloosterboer, N., Tarchini, R., Fiers, M., Sandbrink, H., Lankhorst, R. K. & Vlak, J. M. (2001). The white spot syndrome virus DNA genome sequence. *Virology* **286**, 7-22.

van Hulten, M. C. W., Tsai, M.-F., Schipper, C. A., Lo, C.-F., Kou, G.-H. & Vlak, J. M. (2000). Analysis of a genomic segment of white spot syndrome virus of shrimp containing ribonucleotide reductase genes and repeat regions. *J. Gen. Virol.* **81**, 307-316.

van Regenmortel, M. H. V., Fauquet, C. M., Bishop, D. H. L., Carstens, E. B., Estes, M. E., Lemon, S. M., Maniloff, J., Mayo, M. A., McGeoch, D. J., Pringle, C. R. & Wickner, R. B. (2000). Virus Taxonomy: Seventh Report of the International Committee on Taxonomy of Viruses, pp. 1162. San Diego, USA: Academic Press.

Van Valen, L. (1973). A new evolutionary law. *Evolutionary Theory* **1**, 1-30.

Vassylyev, D. G. & Morikawa, K. (1996). Precluding uracyl from DNA. *Curr. Biol.* **4**, 1381-1385.

Vida, M. G. (1527). De Bombyce. Cremona.

Villareal, L. P. & DeFilippis, V. R. (2000). A hypothesis for DNA viruses as the origin of eukaryotic replication proteins. *J. Virol.* **74**, 7079-7084.

Volkman, L., Blissard, G. W., Friesen, P., Keddie, B. A., Posse, R. & Theilmann, D. A. (1995). Family *Baculoviridae.* In *Virus Taxonomy. Sixth Report of the International Committee on Taxonomy of Viruses*, pp. 104-113. Edited by F. A. Murphy, C. M. Fauquet, D. H. L. Bishop, S. A. Ghabrial, A. W. Jarvis, G. P. Martelli & M. D. Summers. Vienna and New York: Springer- Verlag.

von Mering, C., Krause, R., Snel, B., Cornell, M., Oliver, S. G., Fields, S. & Bork, P. (2002). Comparative assessment of large-scale data sets of protein-protein interactions. *Nature* **417**, 399-403.

Wahlberg, N. (2001). The phylogenetics and biochemistry of host-plant specialization in Melitaeinae butterflies (Lepidoptera: Nymphalidae). *Evolution* **55**, 522-537.

Wang, P. & Granados, R. R. (2000). Calcofluor disrupts the midgut defense system in insects. *Insect Biochem. Mol. Biol.* **30**, 135-143.

Washburn, J. O., Wong, J. F. & Volkman, L. E. (2001). Comparative pathogenesis of Helicoverpa zea S nucleopolyhedrovirus in noctuid larvae. *J. Gen. Virol.* **82**, 1777-1784.

Westenberg, M., Wang, H. L., Ijkel, W. F. J., Goldbach, R. W., Vlak, J. M. & Zuidema, D. (2002). Furin is involved in baculovirus envelope fusion protein activation. *J. Virol.* **76**, 178-184.

Whelan, S., Liò, P. & Goldman, N. (2001). Molecular phylogenetics: state-of-the-art methods for looking into the past. *Trends Genet.* **17**, 262-272.

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

Whitford, M., Stewart, S., Kuzio, J. & Faulkner, P. (1989). Identification and sequence analysis of a gene encoding gp67, an abundant envelope glycoprotein of the baculovirus *Autographa californica* nuclear polyhedrosis virus. *J. Virol.* **63**, 1393-1399.

Whiting, M. F. (2002). Phylogeny of the holometabolous insect orders based on 18S ribosomal DNA: when bad things happen to good data. In *Molecular*

Systematics and Evolution: Theory and Practice, pp. 69-84. Edited by R. DeSalle, G. Giribet & W. Wheeler. Basel - Boston - Berlin: Birkhauser Verlag.

Wiley, E. O. (1981). Phylogenetics: the theory and practice of phylogenetic systematics. Edited by E. O. Wiley. New York: Wiley.

Williams, G. V. & Faulkner, P. (1997). Cytological Changes and Viral Morphogenesis during Baculovirus Infection. In *The Baculoviruses*, pp. 61-107. Edited by L. K. Miller. New York: Plenum Press.

Wilson, M. E., Mainprize, T. H., Friesen, P. D. & Miller, L. K. (1987). Location, transcription and sequence of a baculovirus gene encoding a small arginine-rich polypeptide. *J. Virol.* **61**, 661-666.

Winstanley, D. & O'Reilly, D. (1999). Granuloviruses. In *Encyclopedia of Virology*, 2nd edn, pp. 140-146. Edited by R. G. Webster & A. Granoff. London: Academic Press.

Wormleaton, S. L. & Winstanley, D. (2001). Phylogenetic analysis of conserved genes within the ecdysteroid UDP-glucosyltransferase gene region of the slow-killing *Adoxophyes orana* granulovirus. *J. Gen. Virol.***82**, 2295-2305.

Yang, S. & Miller, L. K. (1999). Activation of baculovirus very late promoters by interaction with very late factor 1. *J. Virol.* **73**, 3404-3409.

Yang, Z. (2000). Phylogenetic Analysis by Maximum Likelihood (PAML), 3.0. edn. University College London, London, England.

Yang, Z. H. (1997). PAML: a program package for phylogenetic analysis by maximum likelihood. *Comput. Appl. Biosci.* **13**, 555-556.

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

V. Appendices

A. Appendix 1: Gene homologies relationships

The genes identity, homology and position within each completely sequenced baculovirus genome were transcribed from the publications describing the sequences anaotations (Afonso et al., 2001, Ahrens et al., 1997, Ayres et al., 1994, Chen et al., 2001, Chen et al., 2002, Gomi et al., 1999, Hashimoto et al., 2000, Hayakawa et al., 1999, Hyink et al., 2002, IJkel et al., 1999, Kuzio et al., 1999, Luque et al., 2001, Pang et al., 2001).

The numbers in the table correspond to the position of the genes as they were published. Genes outlined in bold italics are the 30 genes common to all 13 baculovirus genomes. Genes outlined in bold are the additional 32 genes common to all lepidopteran baculoviruses

AcMNPV	BmNPV	OpMNPV	EppoNPV	LdMNPV	SeMNPV	SpltNPV	HaSNPV	HzSNPV	PxGV	XcGV	CpGV	CuniNPV	Gene Name
1	130	10	7										ptp1
3		136		149						127			ctl1
4	133	8	6	29									ac4
5	134	7	5										ac5
6	***135***	***6***	***4***	***137***	***12***	***114***	***117***	***121***	***32***	***35***	***41***	***25***	***lef-2***
7													ac7
8	**1**	**3**	**1**	**1**	**1**	**1**	**1**	**1**	**4**	**1**	**1**		**polh/grn**
9	2	2	136	2	2	2	2	2					orf1629
10	**3**	**1**	**135**	**3**	**3**	**3**	**3**	**3**	**6**	**3**	**3**		**pk1**
11	4	11	8	35									ac11
12				151									ac12
13	**5**	**12**	**10**	**122**	**13**	**128**	**123**	**127**	**54**	**81**	**73**		**38.7K**
14	***6***	***13***	***11***	***123***	***14***	***129***	***124***	***128***	***55***	***82***	***74***	***45***	***lef-1***
15	7	14	12	125	27	121	126	130	118		141		egt
16	8	15	13										odv-e26
17	9	16	14	128	29	119	128	132					ac17
18	10	17	15	158	43	113							ac18
19	11	18	16	159	42	112	115	118					ac19
21	12	19	17	118	34	134	131	135					arif1
22	***13***	***20***	***18***	***119***	***35***	***135***	***132***	***136***	***37***	***45***	***48***	***38***	***ac22***
23	***14***	***21***	***19***	***130***	***8***	***136***	***133***	***137***	***26***	***27***	***31***	***104***	***ac23/ld130***

24	15	44	36	110	32	133	130	134					pkip
25	**16**	**43**	**35**	**37**	**126**	**30**	**25**	**24**	**61**	**89**	**81**		**dbp1**
26	17	42	34	36	125	31	26	25					ac26
27	18	41	33										iap1
28	**19**	**40**	**32**	**38**	**127**	**29**	**24**	**23**	**60**	**88**	**80**		**lef6**
29	**20**	**39**	**31**	**39**	**128**	**28**	**23**	**22**	**17**	**16**	**19**		**ac29**
30	21	38	30										ac30
31	23	29		145	48		106	109	47	68	59		sod
32	**24**	**27**	**25**	**156**	**38**	**122**	**113**	**116**	**104**	**144**	**123**		**fgf**
33					54a								ac33
34	25	26	24	42	124	18	27	27					ac34
35	**26**	**25**	**23**	**43**	**123**	**32**	**28**	**28**	**42**	**52**	**54**		**vubi**
36	**27**	**24**	**22**	**44**	**120**	**33**	**31**	**31**	**45**	**55**	**57**		**39K/pp31**
37	**28**	**23**	**21**	**45**	**119**	**34**	**32**	**32**	**46**	**56**	**58**		**lef11**
38	**29**	**22**	**20**	**46**	**118**	**35**	**33**	**33**	**52**	**79**	**69**		**ac38**
39	30												p43
40	**31**	**45**	**37**	**48**	**115**	**36**	**35**	**35**	**51**	**78**	**68**	**73**	**p47**
41	32	46	38			37	36	36					lef12
42	33	47	39										gta
43	34	48	40		113		37	37					ac43
44	35	49	41		84								ac44
45	36												ac45
46	**37**	**50**	**42**	**131**	**57**	**98**	**96**	**99**	**30**	**149**	**37**		**odv-e66**
47	38	51											ets
48		52	44										etm
49		53											pcna
50	***39***	***54***	***45***	***51***	***112***	***38***	***38***	***38***	***109***	***148***	***131***	***26***	***lef8***
51	40	55	46		111	39	39	39					ac51
52	41			53	109	44	42	43					ac52
53	**42**	**56**	**47**	**54**	**108**	**45**	**43**	**44**	**112**	**171**	**134**		**ac53**
54	***43***	***58***	***49***	***57***	***105***	***49***	***47***	***48***	***115***	***175***	***138***	***8***	***vp1054***
55	44	59	50	58	104	50	48	49					ac55
56	45	60	51		103		49	50					ac56
57	46	61	52	60	102	51	50	51					ac57
58													ac58
59	47	62	53	61	101	52	51	52					ac59
60	48	63	54	62	100	53	52	53		102			ac60
61	**49**	**64**	**55**	**63**	**98**	**57**	**53**	**54**	**100**	**140**	**118**		**fp/25K**
62	***50***	***65***	***56***	***64***	***97***	***59***	***55***	***56***	***99***	***139***	***117***	***59***	***lef9***
63	51			117			121	125					ac63

64	52	69	57	68	25	32.5	58	59		107	13		gp37
65	***53***	***70***	***58***	***83***	***93***	***69***	***67***	***69***	***93***	***132***	***111***	***91***	***dnapol***
66	**54**	**71**	**59**	**82**	**92**	**68**	**66**	**68**	**94**	**133**	**112**		**ac66/desmplakin**
67	**55**	**72**	**60**	**81**	**91**	**67**	**65**	**67**	**95**	**134**	**113**		**lef-3**
68	***56***	***73***	***61***	***80***	***90***	***66***	***64***	***66***	***96***	***135***	***114***	***58***	***ac68***
69	57		69		89	65	63	65					ac69
70													ac70
71	58	74	63	79	88	64	62	64					iap2
72	58a	75	64										ac72
73	59	76	65										ac73
74	60	77	66				68	70					ac74
75	**61**	**78**	**67**	**84**	**94**	**72**	**69**	**71**	**92**	**126**	**108**		**ac75**
76	**62**	**79**	**68**	**85**	**95**	**73**	**70**	**72**	**91**	**125**	**107**		**ac76**
77	***63***	***80***	***69***	***86***	***82***	***74***	***71***	***73***	***89***	***123***	***106***	***18***	***vlf1***
78	**64**	**81**	**70**	**87**	**81**	**75**	**72**	**74**	**88**	**122**	**105**		**ac78**
79	65	82	71							75	65		11.3K
80	***66***	***83***	***72***	***88***	***80***	***76***	***73***	***75***	***87***	***121***	***104***	***33***	***gp41***
81	***67***	***84***	***73***	***89***	***79***	***77***	***74***	***76***	***86***	***120***	***103***	***106***	***ac81***
82	**68**	**85**	**74**	**90**	**78**	**78**	**75**	**77**	**85**	**119**	**102**		**ac82/tlp20**
83	***69***	***86***	***75***	***91***	***77***	***79***	***76***	***78***	***84***	***118***	***101***	***35***	***p95/vp91***
84													ac84
85		87											ac85
86													pnk
87	70	88											vp80/p87
88	71	89	76		76	80	77	80					cg30
89	***72***	***90***	***77***	***92***	***75***	***81***	***78***	***81***	***79***	***111***	***96***	***24***	***vp39***
90	***73***	***91***	***78***	***93***	***74***	***82***	***79***	***82***	***78***	***110***	***95***	***96***	***lef-4***
91	74	92	79						73				ac91
92	***75***	***93***	***80***	***94***	***73***	***83***	***80***	***83***	***76***	***101***	***93***	***14***	***ac92***
93	**76**	**94**	**81**	**95**	**72**	**84**	**81**	**84**	**75**	**100**	**92**		**ac93**
94	**77**	**95**	**82**	**96**	**71**	**85**	**82**	**85**	**74**	**99**	**91**		**p25/odv-e25**
95	***78***	***96***	***83***	***97***	***70***	***86***	***84***	***87***	***72***	***98***	***90***	***90***	***helicase***
96	***79***	***97***	***84***	***98***	***69***	***87***	***85***	***88***	***71***	***97***	***89***	***89***	***ac96***
97													ac97
98	***82***	***99***	***85***	***99***	***67***	***88***	***86***	***89***	***70***	***96***	***88***	***87***	***38K***
99	***83***	***100***	***86***	***100***	***66***	***89***	***87***	***90***	***69***	***95***	***87***	***88***	***lef-5***
100	***84***	***101***	***87***	***101***	***65***	***90***	***88***	***91***	***67***	***94***	***86***	***23***	***p6.9/dnabinding***
101	**85**	**102**	**88**	**102**	**64**	**91**	**89**	**92**	**66**	**93**	**85**		**c42/p40**
102	**86**	**103**	**89**	**103**	**63**	**92**	**90**	**93**	**64**	**92**	**84**		**p12**
103	**87**	**104**	**90**	**104**	**62**	**93**	**91**	**94**	**63**	**91**	**83**		**p45**

104	88	105	91	105	61	94	92	95					vp80/p87
105	89						61	63		67			he65
106	**90**	**107**	**93**	**140**	**53**	**101**	**101**	**104**	**40**	**50**	**52**		**ac106**
107													ac107
108	91	108	94	108	58	97	95	98					ac108
109	***92***	***109***	***95***	***107***	***59***	***96***	***94***	***97***	***43***	***53***	***55***	***69***	***ac109***
110	*92a*	*111*	*96*	*106*	*60*	*95*	*93*	*96*	*41*	*51*	*53*		*ac110*
111	93	112	97	76			116	120		160			ac111
112													ac112
113										147			ac113
114	94	114	100										ac114
115	***95***	***115***	***102***	***143***	***50***	***107***	***98***	***101***	***29***	***32***	***35***	***46***	***ac115***
116	95a												ac116
117	96	117	104		47		110	113					ac117
118													ac118
119	***97***	***119***	***106***	***155***	***36***	***124***	***111***	***114***	***7***	***84***	***75***	***29***	***ac119***
120	98	120	107		37	123							ac120
121	98a												ac121
122	99	121											ac122
123	100												pk2
124	101	122	108										ac124
125	102	123	109										lef7
126	103	124	110	70	19	42	41	41		103	10		chitinase
127	104	125	111	78	16	54	56	57		58	11		vcath
128	105	126	112										gp64/67
129	106	127	114		10	116	118	122	53	80	71		p24
130	107	128	115		9		119	123					gp16
131	108	129	116	136	46	132	120	124					pp34/pep
132	109	130	117										ac132
133	***110***	***131***	***118***	***157***	***41***	***109***	***114***	***117***	***106***	***145***	***125***	***54***	***alk-exo***
134					99					21			p94
135	112					55						75	p35
136	113	132	119	40	129		22	21					p26
137	114	133	120	41	130	19	21	20	2	5			p10
138	***115***	***134***	***121***	***27***	***131***	***21***	***20***	***19***	***49***	***77***	***60***	***74***	***p74***
139	**116**	**137**	**122**	**23**	**7**	**27**	**16**	**16**	**120**	**180**	**143**		**me53**
140													ac140
141	117	138	123	21	138	8	8	8	15	14			ie0
142	***118***	***139***	***124***	***20***	***137***	***11***	***9***	***9***	***14***	***13***	***15***	***30***	***ac142***
143	**119**	**140**	**125**	**19**	**136**	**12**	**10**	**10**	**13**	**12**	**14**		**odv-e18**

144	***120***	***141***	***126***	***18***	***135***	***13***	***11***	***11***	***80***	***112***	***97***	***32***	***odv-ec27***
145	**121**	**142**	**127**	**17**	**134**	**14**	**12**	**12**	**12**	**11**	**9**		**ac145**
146	**122**	**144**	**128**	**16**	**133**	**15**	**13**	**13**	**11**	**10**	**8**		**ac146**
147	**123**	**145**	**129**	**15**	**132**	**16**	**14**	**14**	**10**	**9**	**7**		**ie1**
148	***124***	***146***	***130***	***14***	***6***	***17***	***15***	***15***	***16***	***15***	***18***	***102***	***odv-e56***
149	125												ac149
150	126	110		30	96	22			59	87	79		ac150
151	127	151	131										ie2
152													ac152
153	128	152	133								24		pe38
154	129												ac154
53a	42a	57	48	56	106	48	46	47		174	137		lef10
				4									ld4 mucin-like
				5									ld5
				6									ld6
				7									ld7 g22
				8									ld8
				9									Ld9
				10									ld10
				11									ld11
				12									ld12
				13									ld13
				22					101	141	120		dna-ligase
				24									ld24
				25									ld25
				26			19	18					ld26
				28									ld28
				31									ld31
				34									ld34
				47									dbp2
				49									ld49
				50		40			107	146	126		helicase2
				52									ld52
				55	107	46	44	45					ld55/se107
				59									ld59
				65						152			enhancin-2
		30		66									ctl2
		143		67									hrf1
				69									ld69
				77									ld77

				109									ld109
				111	33								ld111
		31a		116	55	10							dutpase
				120	45	63							rr2b
				121									ld121
				124	15	130							ld124
				126									ld126
				127	28								ld127
				129	30	118	129	133					ld129
				132									ld132
				133									ld133
				134									ld134
				135		71							ld135
				137a	11								orf4 PE
		31		138	54	100					16		ld138
		35	26	139	110	64	103	107			17		iap3
				141	52	102	100	103					parg/ld141
				142	51								ld142
				144	49								ld144
		34		147							128		rr2 (a)
		32		148	139	23					127		rr1
				152									ld152
				160						154			enhancin-3
				162									ld162
				163									ld163
									5	2	2		xc2
										4			xc4
										6			xc6
									8	7	4		xc7
									9	8	5		xc8
									20	17	20		xc17
									23	18	23		xc18
									21	19	22		xc19
										20			xc20
										22			xc22
										23			xc23
										24			xc24
									24	25	29		xc25
									25	26	30		xc26
										28			xc28

									28	29	33		xc29
			101							30	34		xc30
			101a							31			xc31
										33	38		xc33
									31	34	39		xc34
										36	42		xc36
										37	45		xc37
										38			xc38
									34	39			xc39
									35	40	46		xc40/mpnase
										41			xc41
										42			xc42
					56	99	97	100	36	43	47		p13/glycogenin
										44			xc43
										46	49		xc46
									38	47	50		xc47
										48			xc48
										49			xc49
									44	54	56		xc54
										57			xc57
										59			xc59
										61			xc61
										62			xc62
										63			xc63
										64			xc64
										65			xc65
										66	64		xc66
										69			xc69
										70			xc70
										71			xc71
										72			xc72
										73			xc73
										74			xc74
							57	58		83			xc83
									56	85	76		xc85
									57	86	77		xc86
									62	90	82		xc90
										104			xc104
										105			xc105
										106			xc106

										108			xc108
									82	113	99		xc113
										114			xc114
										115			xc115
									83	116	100		xc116
										117			xc117
									90	124			xc124
					17					128			xc128
					21					129			xc129
									97	136	115		xc136
									98	137	116		iap5
										138			xc138
									102	142	121		xc142
									103	143	122		xc143
										150			enhancin-1
										151			xc151
										153			xc153
										155			xc155
										156			xc156
										157			xc157
										158			xc158
										161			xc161
										162			xc162
										163			xc163
										164			xc164
									110	165	132		xc165
										166			enhancin-4
										167			xc167
										168			xc168
									65	169	119		xc169
										170	133		xc170
									113	172	135		xc172
									114	173	136		xc173
									116	176			xc176
										177			xc177
									117	178	140		xc178
										179			xc179
										181			xc181
					18					128a			xc128a
					4	4	4	4					hoar

						5							sl5
						6							sl6
						7							sl7
						9							sl9
						20							sl20
						24							sl24
						25							sl25
						26							sl26
						41							sl41
						43							sl43
						47							sl47
						56							sl56
						58							sl58
						60							sl60
						61							sl61
						62							sl62
						70							sl70
						103							sl103
						104							sl104
					22	105							se22
					23	105a							se23
					24	105b							se24
						106							sl106
						108							sl108
		98				110							op98
						111							sl111
						115							sl115
						117							sl117
						126							sl126
						127							sl127
						131							sl131
						137							sl137
						138							sl138
						139							sl139
						140							sl140
						141							sl141
									1				px1
									3				px3
									18				px18
									19				px19

									22				px22
									27				px27
									33				px33
									39				px39
									48				px48
									50		62		cp62
									58				px58
									68				px68(ac145?)
									77				px77
									81				px81
									105				px105
									108		130		cp130
									111				px111
									119				px119
	111												bm111
												10	cn10
												100	cn100
												101	cn101
												103	cn103
												105	cn105
												107	cn107
												11	cn11
												12	cn12
												13	cn13
												15	cn15
												16	cn16
												17	cn17
												19	cn19
												2	cn2
												20	cn20
												21	cn21
												22	cn22
												27	cn27
												28	cn28
												3	cn3
												31	cn31
												34	cn34
												36	cn36
												37	cn37
												39	cn39

40	cn40
41	cn41
42	cn42
43	cn43
44	cn44
47	cn47
48	cn48
49	cn49
50	cn50
51	cn51
52	cn52
53	cn53
55	cn55
56	cn56
57	cn57
6	cn6
60	cn60
61	cn61
62	cn62
63	cn63
64	cn64
65	cn65
66	cn66
67	cn67
68	cn68
7	cn7
70	cn70
71	cn71
72	cn72
76	cn76
77	cn77
78	cn78
79	cn79
80	cn80
81	cn81
82	cn82
83	cn83
84	cn84
85	cn85/obp
86	cn86

												9	cn9
												92	cn92
												93	cn93
												94	cn94
												97	cn97
												98	cn98
												99	cn99
											100		
											109		cp109
											110		cp110
											12		cp12
											124		cp124
											129		cp129
											139		cp139
											142		cp142
											21		cp21
											25		cp25
											26		cp26
											27		cp27
											28		cp28
											32		cp32
											36		cp36
											40		cp40
											43		cp43
											44		cp44
											51		cp51
											6		cp6
											61		cp61
											67		cp67
											70		cp70
											72		cp72
											78		cp78
			105										ep105
			113										ep113
			132										ep132
			134										ep134
			28										ep28
			43										ep43
			9										ep9
			98										ep98
							102	106					ha102

							104	108						ha104
							107	110						ha107
							108	111						ha108
							109	112						ha109
							112	115						ha112
							122	126						ha122
							125	129						ha125
							127	131						ha127
							134	138						ha134
							135	139						ha135
							18	17						ha18
							29	29						ha29
							30	30						ha30
							34	34						ha34
							40	40						ha40
							45	46						ha45
							5	5						ha5
							54	55						ha54
							6	6						ha6
							7	7						ha7
							99	102						ha99
							115a	119						hz119
								26						hz26
								42						hz42
								62						hz62
								79						hz79
											94			iap cp1
		106	92											iap op4
					114									odv-e66(2)
		113	99											op113
		118												op118
		135												op135
		150												op150
		28												op28
		33												op33
		36	27											op36
		37	29											op37
		4	2											op4
		5	3											op5
		66												op66

		67											op67
		68											op68
		148											opep25
		147											opep32
					87								p26(2)
		149											p8.9
		9			26						66		ptp2
											98		ptp2a
					116								se116
					117								se117
					121								se121
					122								se122
					20								se20
					31								se31
					39								se39
					40								se40
					44								se44
					5								se5
					68		83	86					se68
					83								se83
					85								se85
					86								se86

B. Appendix 2: Virus collection

The virus isolates that were available for DNA sequencing are listed in the table bellow. When no other indication was available, isolates were referred as belonging to the IVEM collection (NERC-CEH Oxford). There is therefore little evidence of original sampling location for most isolates.

Order/ Lepidopteran Family	Species name	Virus type	Acc#	Collector	Locality	Country
Coleoptera	*Oryctes rhinoceros*	Oryctes	576	I.A.D. Robertson		Seychelles
Diptera	*Tachinid flies*	NOV	770		IVEM Oxford	UK
Diptera	*Tipula oleracea*	NPV	35	K.M. Smith	IVEM Oxford	UK
Diptera	*Tipula oleracea*	NPV	281	K.M. Smith	IVEM Oxford	UK
Diptera	*Tipula paludosa*	IV	419		IVEM Oxford	UK
Hymenoptera	*Anoplonyx destructor*	NPV	177	J.F. Longworth	IVEM Oxford	UK
Hymenoptera	*Gilpinia hercyniae*	NPV	239	P.F. Entwistle	IVEM Oxford	UK
Hymenoptera	*Gilpinia hercyniae*	NPV	i 7		IVEM Oxford	UK
Hymenoptera	*Gilpinia hercyniae*	NPV	K14		IVEM Oxford	UK
Hymenoptera	*Nematus olfasciens*	NPV	76	K.M. Smith	IVEM Oxford	UK
Hymenoptera	*Neodipirion lecontei*	NPV	726	Cunningham	FPMI	Canada
Hymenoptera	*Neodipirion sertifer*	NPV	345		IVEM Oxford	UK
Hymenoptera	*Neodipirion sertifer*	NPV	411		IVEM Oxford	UK
Hymenoptera	*Neodipirion sertifer*	NPV	413		Scotland	UK
Hymenoptera	*Neodiprion sertifer*	NPV	k5		IVEM Oxford	UK
Hymenoptera	*Neodiprion sertifer*	NPV	M1		IVEM Oxford	UK
Neuroptera	*Chrysopa sp*	PV!	330	K.M. Smith	IVEM Oxford	UK
Neuroptera	*Hemerobius sp*	?PV	404	K.M. Smith	IVEM Oxford	UK
Neuroptera	*Hemerobius sp*	NPV	318	K.M. Smith	IVEM Oxford	UK
Arctiidae	*Ammalo insulata*	NPV	175		IVEM Oxford	UK
Arctiidae	*Diacrisia obliqua*	NPV	798	Dr Pawar		India
Arctiidae	*Euplagia quadripunctaria*	NPV	387	K.M. Smith	IVEM Oxford	UK
Arctiidae	*Hyphantria cunea*	GV	228		IVEM Oxford	UK
Arctiidae	*Hyphantria cunea*	NPV	N3			Japan
Arctiidae	*Spilosoma imparilis*	NPV	N2			Japan
Arctiidae	*Spilosoma lutea*	NPV	126	K.M. Smith	IVEM Oxford	UK
Arctiidae	*Tyria jacobaeae*	NPV+CPV	223		IVEM Oxford	UK
Arctiidae	*Utetheisa pulchella*	NPV+CPV	142	K.M. Smith	IVEM Oxford	UK
Bombycidae	*Bombyx mori*	NPV	460	P. Seraphim		Poland
Gelechiidae	*Phthorimaea operculella*	GV	c3	J. Cory	isolate 8	Nicaragua
Gelechiidae	*Phthorimaea operculella*	GV	c4	J. Cory	isolate 9	Nicaragua
Geometridae	*Abraxas grossulariata*	NPV	112	K.M. Smith	IVEM Oxford	UK
Geometridae	*Buzura edwardsii*	NPV	505	A.J. Weaving		Rodhesia
Geometridae	*Buzura edwardsii*	NPV	504	A.J. Weaving		Rodhesia
Geometridae	*Erannis (Hibernia) defoliaria*	NPV	174		IVEM Oxford	UK

Geometridae	*Idaea virgularia*	NPV	402	K.M. Smith	IVEM Oxford	UK
Geometridae	*Lambdina fiscellaria lugubrosa*	NPV	N5			Canada
Geometridae	*Operopthera brumata*	NPV	291	J. Cory	Dartmoor	UK
Geometridae	*Operopthera brumata*	NPV	i 4	J. Cory	Lanarh	UK
Geometridae	*Operopthera brumata*	NPV	i 5	J. Cory	Wytham Wood	UK
Geometridae	*Rhodometra sacraria*	NPV	339	K.M. Smith	IVEM Oxford	UK
Glyphipterigidae	*Anthophila fabricana*	GV	790	Dr Shaw	Reading Univ.	UK
Hepialidae	*Wiseana cervinata*	GV	342			New Zeland
Hepialidae	*Wiseana cervinata*	NPV	344			New Zeland
Lasiocampidae	*Lasiocampa quercus*	NPV	10		IVEM Oxford	UK
Lasiocampidae	*Lasiocampa quercus*	NPV	202		IVEM Oxford	UK
Lasiocampidae	*Malacosoma californicum pluviale*	NPV	k3	J. Myers		Canada
Lasiocampidae	*Malacosoma californicum pluviale*	NPV	k13	J. Myers		Canada
Lasiocampidae	*Malacosoma disstria*	NPV	N7	J. Myers		Canada
Lasiocampidae	*Malacosoma neustria*	NPV	k12	J. Myers		Canada
Lasiocampidae	*Pachypasa papyri*	NPV	436	Mrs Aloo		Kenya
Lasiocampidae	*Pachypasa papyri*	NPV	436d8	Mrs Aloo		Kenya
Lasiocampidae	*Trabala vishnou*	NPV	195	J. F. Longworth	IVEM Oxford	UK
Limacodidae	*Darna trima*	GV	545	Dr Munroe	Sarawak	Malaysia
Limacodidae	*Natada nararia*	GV	254	K.M. Smith	IVEM Oxford	UK
Lymantriidae	*Aroa discalis*	NPV	63	K.M. Smith		Nigeria
Lymantriidae	*Euproctis chrysorrohea*	NPV	a10	J. Cory	IVEM Oxford	UK
Lymantriidae	*Euproctis similis*	NPV	768	H. Evans	Kent	UK
Lymantriidae	*Lymantria monacha*	NPV	b4	J. Cory	IVEM Oxford	UK
Lymantriidae	*Lymantria ninayi*	NPV	750	Dr Hywal Roberts		New Guinea
Lymantriidae	*Lymantria ninayi*	NPV	743	Dr Hywal Roberts		New Guinea
Lymantriidae	*Lymantria obfuscata*	NPV	193			Kashmir
Lymantriidae	*Orgyia mixta*	NPV	67	K.M. Smith		Nigeria
Lymantriidae	*Orgyia antiqua*	NPV	a9	J. Cory	IVEM Oxford	UK
Lymantriidae	*Orgyia leusostigma*	SNPV	b6	J. Cory	IVEM Oxford	UK
Lymantriidae	*Orgyia pseudotsugata*	NPV	b5	J. Cory	IVEM Oxford	UK
Lymantriidae	*Orgyia pseudotsugata*	NPV	N6	J. Cory		Canada
Lymantriidae	*Stilpnotia salicis*	?PV	370	Dr Jaroslav Weiser		
Noctuidae	*Achaea faber*	NPV	64	K.M. Smith		Nigeria
Noctuidae	*Achaea janata*	GV	835	Dr Pawar		India
Noctuidae	*Agrotis segetum*	NPV	i 6	J. Cory	IVEM Oxford	UK
Noctuidae	*Agrotis segetum*	NPV	k4	J. Cory	IVEM Oxford	UK
Noctuidae	*Anagrapha falcifera*	NPV	k2	Hoffman	1998	USA
Noctuidae	*Anomis sabulifera*	NPV	14	K.M. Smith	IVEM Oxford	UK
Noctuidae	*Anticarsia gemmatalis*	NPV	b10	J. Cory	IVEM Oxford	UK
Noctuidae	*Autographa californica*	NPV	k1	Hoffman	1998	USA
Noctuidae	*Cerapteryx graminis*	NPV	a5		IVEM Oxford	UK
Noctuidae	*Cirphis unipuncta*	PV	172	K.M. Smith	IVEM Oxford	UK
Noctuidae	*Euplexia lucipara*	GV	248	K.M. Smith	IVEM Oxford	UK
Noctuidae	*lacanobia oleracea*	GV	a6	I. Smith		UK
Noctuidae	*Lithophane leautieri*	NPV+CPV	134	K.M. Smith	IVEM Oxford	UK
Noctuidae	*Mamestra brassicae*	NPV	274		IVEM Oxford	UK
Noctuidae	*Mamestra brassicae*	NPV	k9	J. Cory	IVEM Oxford	UK
Noctuidae	*Melanchra persicariae*	GV	26	K.M. Smith	IVEM Oxford	UK
Noctuidae	*Mythimna (Pseudolatia) unipuntata*	NPV	a1	J. Cory		USA
Noctuidae	*Mythimna separata*	NPV	a2	J. Cory		Japan

Noctuidae	*Mythimna separata*	NPV	a3	J. Cory		China
Noctuidae	*Mythimna separata*	NPV	a4	J. Cory		India
Noctuidae	*Panolis flammea*	NPV	i 3	D. Hdgson	Scotland	UK
Noctuidae	*Plusia gamma*	NPV	287		IVEM Oxford	UK
Noctuidae	*Spodoptera exempta*	NPV	k11	J. Cory	IVEM Oxford	UK
Noctuidae	*Spodoptera frugiperda*	NPV	k10	J. Cory	IVEM Oxford	UK
Noctuidae	*Spodoptera littoralis*	GV	66			Sri Lanka
Noctuidae	*Spodoptera littoralis*	NPV	440	J. Cory	IVEM Oxford	UK
Noctuidae	*Spodoptera litura*	GV	766	Dr T. Sankaran	Bangalore	Sri Lanka
Noctuidae	*Spodoptera sunia*	NPV	k7	J. Cory	unan 13	Nicaragua
Noctuidae	*Spodoptera sunia*	NPV	k8	J. Cory	unan 10	Nicaragua
Noctuidae	*Trichoplusia ni*	MNPV	c2	J. Cory	IVEM Oxford	UK
Noctuidae	*Trichoplusia orichalcea*	NPV	b9	J. Cory	IVEM Oxford	UK
Notodontidae	*Cerura vinula*	?NPV	313	K.M. Smith	IVEM Oxford	UK
Notodontidae	*Cynosarga chrysolopa*	?NPV	302	K.M. Smith	IVEM Oxford	UK
Notodontidae	*Phalera bucephala*	NPV+CPV	204		IVEM Oxford	UK
Nymphalidae	*Aglais urticae*	NPV	a7	J. Cory	IVEM Oxford	UK
Nymphalidae	*Agraulis vanillae*	NPV	779	D. Rushen	Dorset	UK
Nymphalidae	*Helioconius erato*	NPV	789	D. Rushen	Worldwide Butterflies	UK
Nymphalidae	*Junonia coenia*	GV	19	K.M. Smith	IVEM Oxford	UK
Nymphalidae	*Nymphalis io*	NPV	353	K.M. Smith	IVEM Oxford	UK
Nymphalidae	*Polygonia c-album*	NPV	380	K.M. Smith	IVEM Oxford	UK
Nymphalidae	*Vanessa atalanta*	NPV	54	K.M. Smith	IVEM Oxford	UK
Nymphalidae	*Vanessa cardui*	?PV	367	K.M. Smith	IVEM Oxford	UK
Pieridae	*Aporia crataegi*	NPV	106	K.M. Smith	IVEM Oxford	UK
Pieridae	*Aporia crataegi*	NPV	125	K.M. Smith	IVEM Oxford	UK
Pieridae	*Catopsila pomona*	NPV	78	K.M. Smith		New Guinea
Pieridae	*Gonepteryx rhamni*	NPV	103	K.M. Smith	IVEM Oxford	UK
Pieridae	*Pieris (Artogeia) napi*	GV	242	K.M. Smith	IVEM Oxford	uK
Pieridae	*Pieris brassicae*	?GV	384	K.M. Smith	IVEM Oxford	UK
Pieridae	*Pieris rapae*	GV	95	K.M. Smith	IVEM Oxford	UK
Psychidae	*Mahasena corbetti*	GV	699	Dr Salman Shah		Malaysia
Psychidae	*Mahasena corbetti*	NPV	692	Dr Salman Shah		Malaysia
Psychidae	*Mahasena corbetti*	NPV	595	Sabah		
Psychidae	*Kotochalia junodi*	NPV	205	K.M. Smith	IVEM Oxford	UK
Psychidae	*Kotochalia junodi*	NPV	32	K.M. Smith	IVEM Oxford	UK
Pyralidae	*Diatrea saccharalis*	NPV	k6	J. Cory	IVEM Oxford	UK
Pyralidae	*Ephestia cautella*	NPV	c5	J. Cory	IVEM Oxford	UK
Pyralidae	*Plodia interpunctella*	GV	b3	J. Cory	IVEM Oxford	UK
Saturniidae	*Actias selene*	NPV	47	K.M. Smith	IVEM Oxford	UK
Saturniidae	*Antheraea (Telea) polyphemus*	NPV	30	K.M. Smith	IVEM Oxford	UK
Saturniidae	*Samia (Philosaria) ricini*	NPV	28	K.M. Smith	IVEM Oxford	UK
Sphyngidae	*Laothoe populi*	NPV	358	K.M. Smith	IVEM Oxford	UK
Sphyngidae	*Smerinthus ocellata*	NPV+CPV	819	R. Watson	Lymington	UK
Sphyngidae	*Sphinx ligustri*	NPV	362	K.M. Smith	IVEM Oxford	UK
Sphyngidae	*Sphinx ligustri*	NPV+CPV	334	K.M. Smith	IVEM Oxford	UK
Sphyngidae	*Sphinx ligustri*	NPV+CPV	400	K.M. Smith	IVEM Oxford	UK
Sphyngidae	*Sphinx ligustri*	NPV+CPV	144	K.M. Smith	IVEM Oxford	UK
Tortricidae	*Cacoecia murinana*	GV	37	Dr Jaroslav Weiser		
Tortricidae	*Homona coffearia*	GV	744	D. Shaw		New Guinea
Tortricidae	*Homona coffearia*	NPV	745	D. Shaw		New Guinea

Zygaenidae	*Harrisina brillans*	GV	M2	B. Federici		USA
	Cynosarga	?NPV	407	K.M. Smith	IVEM Oxford	UK

C. Appendix 3: *polyhedrin* and *lef-8* DNA sequence alignments

Bellow arethe *polyhedrin* and *lef-8* alignments used for the phylogenetic analyses in the lepidopteran baculovirus studies.

polh Alignment

```
Abraxasgrossulariata_NPV112     TTCGTGGGGC CGGGGAAGAA CCAAAAGCTG ACCCTGTTTA AAGAAATCCG CAACATCAAG CCCGACACGA TGAAACTCAT
Achaea_faber_NPV64              TTCGTGGGGC CGGGGAAGAA TCAGAAGCTG ACGCTTTTCA AAGAAATCCG TAACGTCAAG CCCGACACGA TGAAGCTCGT
AchaeajanataGV835               GTTGTGGGGC CGGGGAAGAA TGTGCGCATA aCCCTGTTGA AGGAAATTAG ACGCATCCAC CCCGACACAA TGAAGTTGGT
AcMNPV                          TTCCTGGGAC CCGGCAAGAA CCAAAAACTC ACTCTCTTCA AGGAAATCCG TAATGTTAAA CCCGACACGA TGAAGCTTGT
AcMNPVK1                        NNNNNNNNNN NNNNNNNCTA CCCAAAACTC ACTCTCTTTA AGGAAATCCG TAATGTTAAA CCCGACACGA TGAAGCTTGT
Actias_selene_NPV47             AGTTGGGGGC CGGGGAAGAC CAAAAAACTG ACACTTTTCA AGGAAATCCG CAACGTTAAA CCCGACACAA TGAAACTTAT
Adoxophyes_orana_GV             TTTCGTGGTC CGGGTAAAAA TGTACGCATC ACCTTGTTCA AAGAAATTCG TCGAGTACAT CCCGATACGA TGAAGCTTGT
AfMNPVK2                        NNNNNNNNNN NNNNNNNNNN NCAAAAACTC ACTCTTTTCA AAGAAATCCG TAATGTTAAA CCCGATACCA TGAAGTTGAT
Aglais_urticae_NPVA7            TTCGTGGGGC CGGGGAAGAA TCAAAAACTG ACATTGTTTA AAGAGATTCG TAATGTCAAA CCCGATACGA TGAAGCTCGT
Agraulis_vanillare_NPV779       TTCGTGGGGC CGGGGAAGAA YCAAAAATTG ACCCTGTTTA AAGAAATTCG CAACGTCAAG CCCGACACGA TGAAACTTAT
Agrotis_segetum_NPVi6           GTCGTGGGGC CGGGGAAGAA TCAGAAACTG ACCCTCTTCA GAGAGATTCG TGATGTAAAA CCCGACACCA TGAAGTTGGT
Agrotis_segetum_NPVK4           NNNNGGGGGC CGGCGAAGAT TCAGAAATTA ACCCTAGTTA GAGAGATTCG TGATGTAAAA CCCGACACCA TGAAGTTGGT
Ammalo_insulata_NPV175          TTCGTGGGGC CGGGGAAGAC CCAAAAACTG ACACTCTTCA AGGAAGTCCG CAACGTCAAG CCCGACACAA TGAAACTCAT
Amonis_sabulifera_NPV14         TTCGTGGGGC CGGGGAAGAA CCAAAAATTA ACTCTGTTCA AGGAGATCCG TAATGTCAAG CCCGACACAA TGAAACTAGT
Amsacta_albistriga_NPV          TTCTTGGGGC CGGGTAAAAA TCAAAAGTTA ACTTTGTTCA AAGAGATTCG TAATGTAAAA CCGGACACGA TGAAACTGAT
Antherea_pernyi_NPV             TTCCTGGGGC CGGGTAAAAA CCAAAAACTG ACACTTTTCA AGGAAATCCG CAACGTTAAA CCCGACACAA TGAAACTTAT
Antherea_polyphemus_NPV30       AGTCGGGGGC CGGGGAAGAA CCAAAAACTT ACCCTTTTTA AAGAGATTCG CAATGTGAAA CCCGATACCA TGAAGTTAAT
Anthophila_fabricana_GV790      TTCGGGGGGC CGGGGAAGAA TGTGAGAATA ACACTGTTTA AAGAGATCAG GCGCGTCCAC CCGGACACAT TGAAACTTGT
Anticarsia_gemnitalis_NPV       TTTTTAGGGC CCGGCAAAAA CCAAAAATTG ACACTTTTTA AAGAAATCCG CAATGTAAAG CCCGACACGA TGAAGCTCAT
Aporia_crataegi_NPV106          TTCGTGGGGC CGGGGAAGAA CCAAAAATTG ACTTTGTTTA AAGAAATCCG CAACATCAAA CCCGACACGA TGAAACTGGT
Archips_cerasivora_NPV          TTTCTCGGCC CGGGCAAGAA CCAAAAGCTG ACCCTGTTTA AAGAAATCCG CAATGTTAAA CCCGACACGA TGAAACTCAT
Aroa_discalis_NPV63             TTCGTGGGGC CGGGGAAGAA CCAAAAACTC ACTCTACTCA AAGAGATCCG CAACGTTAAG CCCGACACAA TGAAACTGGT
Artogeia_napi_GV242             GTTGTGGGGC CGGGGAAGAA TGTTAGGATT ACCCTTTTCA AAGAATGTAG ACGAGTCGAA CCCGACACAT TGAAGCTGGT
Attacus_ricini_NPV              TTCCTGGGGC CGGGTAAAAA CCAAAAACTG ACACTTTTCA AGGAAATCCG CAACGTTAAG CCTGACACAA TGAAACTTAT
BmNPV                           TTTTTAGGAC CGGGCAAAAA CCAAAAACTT ACCCTTTTTA AAGAAATTCG CAGTGTGAAA CCCGATACCA TGAAGTTAAT
Bombyx_mori_NPV460              NCAGTGGGGC CGGGGAAGAA CCAAAAACTT ACCCTTTTTA AAGAGGTTCG CAATGTGAAA CCCGATACCA TGAAGTTAAT
Buzura_edwardsii_NPV505         TTCGTGGGGC CGCCGAAGAA CCAAAAGCTG ACTCTGTTCA AAGAGATCTG TATTGTCAAG CCCGACACCA TGAAGCTGGT
Buzura_Supressaria_NPV          TTTCTTGGAC CCGGTAAAAA CCAAAAACTT ACTTTGTTTA AAGAAATTCG TAATGTTAAA CCAGACACCA TGAAATTGGT
Cacoecia_murinana_GV37          TCGTGGGGGC CGGGGAAGAA CGTACGAATT ACCCTTTTCA AGGAAATTAG GCGCGTTCAA CCAGACACCA TGAAATTGGT
Catopsila_pomona_NPV78          TTCGTGGGGC CGGGGAAGAA TCAAAAATTA ACATTGTTCA AAGAAATTCG CAATGTGAAG CCCGACACCA TGAAACTTGT
Cerapteryx_gramini_NPVA5        GTCGTGGGGC CGGGGAAGAA CCAAAAGCTG ACCCTGTTCA AAGAGATTCG YATTGTCAAG CCCGACACCA TGAAACTGAT
ChfuNPV                         TTCTCGGGTC CTGGCAAAAA CCAAAAACTA ACTCTCTTCA AAGAGATTCG CAATGTTAAG CCCGACACGA TGAAACTCAT
ChfuNPVDEF                      TTCTTAGGAC CCGGCAAAAA CCAAAAACTA ACCCTTTTTA AAGAAATCCG CAACGTAAAA CCCGACACGA TGAAGCTCAT
Choristoneura_furamifera_GV     TTTCGTGGAC CCGGTAAGAA CGTACGAATT ACCCTTTTCA AGGAAATTAG GCGCGTTCAA CCAGACACCA TGAAATTGGT
Choristoneura_rosaceana_NPV     TTTCTCGGTC CTGGCAAAAA CCAAAAACTA ACTCTCTTCA AAGAGATTCG CAATGTTAAA CCCGACACGA TGAAACTCAT
Cirphis_unipuncta_NPV172        TTCGGGGGGC CGGCGAAGAA CCAAAAGCCA ACTCTCCTCA AAGAGATCAG TAATGTTAAG CCCGACACGA TGAAGCTGGT
CpGV                            TTTCGTGGGC CCGGCAAGAA TGTAAGAATC ACCCTGTTCA AGGAGATTAG ACGCGTCCAC CCAGACACAA TGAAGCTGGT
Cryptophlebia_leucotreta_GV     TTTCGTGGAC CCGGCAAGAA TGTAAGGATC ACCCTTTTCA AGGAAATTAG ACGTGTCCAC CCAGACACAA TGAAGCTGGT
Cynosarga_chrysolopa_NPV302     TTCGTGGGGC CGGGGAAGAA CCAAAAATTG ACACTGTTCA AGGAAATTCG CAATGTAAAG CCCGACACTA TGAAATTAAT
Darna_trima_GV545               TTCGTGGGGC CGGGGAAGAA TGTGAGGATT ACGCTTTTTA AAGAAATTAG GCGCGTACAC CCCGACACCA TGAAACTGGT
Diacrisia_obliqua_NPV798        TTCGTGGGGC CGGGGAAGAA CCAAAAATTG ACGCTTTTCA AAGAAATCCG CAACGTCAAG CCCGACACAA TGAAACTGAT
Diatrea_saccharalis_NPVK6       TTCGGGGGGC CGGGGAAGAA CGTTAGAATA ACATTACTTA AAGAGATTAG ACGCATTCAA CCAGATACAA TGAAACTTGT
Ectropis_obliqua_NPV            TTCATGGGGC CGGGCAAGAA TCAAAAACTA ACTTTGTTCA AAGAGATTCG CAACGTCAAG CCCGACACGA TGAAACTGAT
Ephestia_cautella_NPVC5         TTCGTGGGGC CGGGGAAGAA CCAAAGATTG ACTCTATTCA AGGAGATTCG CAATGTCAAG CCTGATACCA TGAAACTCAT
EppoMNPV                        TTCCTTGGGC CGGGCAAAAA TCAAAAATTA ACCCTATTTA AAGAAATCCG CAACGTCAAG CCCGACACGA TGAAACTCAT
Erannis_defolariaNPV174         TTCGTGGGGC CGGGGAAGAA TCAAAAATTG ACTCTTTTCA AGGAGATTCG TAATATTAAA CCAGACACAA TGAAACTTGT
Euplasia_luciparaGV248          GCAGTGGGGC CGGGGAAGAA CGTGAGAATT ACGCTCTTCA AGGAAATCCG TAGGGTTCAG CCCGACACCA TGAAACTCAT
Euproctis_Chrysorrohea_NPVA10   TTCGTGGGGC CGGGGAAGAA CCAAAAACTC ACACTGTTCA AAGAGATTCG TAACGTAAAA CCCGACACTA TGAAATTGAT
Euproctis_similis_NPV768        TTCGTgGGGC CGGGGAAGAA CCAAAAACTC ACACTGTTCA AAGAGATTCG TAACGTAAAA CCCGACACTA TGAAATTGAT
Harrisina_brillians_GV          TTTCGTGGCC CTGGTAAAAA TGTTAGAATA ACATTATTCA AAGAAATTAG GCGTGTTCAA CCAGACACAA TGAAGTTAGT
Harrisina_brillians_GVM2        TTCGGGGGGC CGGGGAAGAA TGTTAGAATA ACATTATTCA AAGAAATTAG GCGTGTTCAA CCAGACACAA TGAAGTTAGT
HaSNPV                          TTTTTGGGAC CTGGCAAAAA TCAAAAACTA ACTTTGTTTA AAGAGATTCG CAGCGTTAAG CCCGACACAA TGAAGCTTGT
Heliconius_erato_NPV789         TTCGTGGGGC CGGGGAAGAA CCAAAAATTG ACCCTGTTTA AAGAAATTCG CAACGTCAAG CCCGACACGA TGAAGTTGAT
Homona_coffearia_GV744          NNNNNNNNNN NNNNNNNNNN ACGTACGATT ACCCTTTTGA AGGAAATTAG GCGCGTTCAA CCAGACACCA TGAAGTTGGT
Homona_coffearia_NPV745         GTCGTGGGGC CGGGGAAGAA CGTACGAATT ACCCTTTTCA AGGAAATTAG GCGCGTTCAA CCAGACACCA TGAAGTTGGT
Hyphantria_cunea_GV228          TTCGTGGGGC CGGGGAAGAA TGTTAGGATC ACTTTGTTCA AAGAGATTCG TCGCGTCCAG CCAGACACTT TGAAATTAGT
Hyphantria_cunea_NPV            TTCTTAGGAC CCGGCAAGAA CCAAAAATTG ACGCTTTTCA AAGAAATCCG CAACGTCAAG CCCGACACAA TGAAACTGAT
Hyphantria_cunea_NPVN3          NNNNNNNNNN NNNNNNNNAA CCAAAAATTG TCGCTTTTTA AAGAAATCCG CAACGTCAAG CCCGACACAA TGAAACTGAT
HzSNPV_genome                   TTTTTGGGAC CTGGCAAAAA TCAAAAACTA ACTTTGTTTA AAGAGATTCG CAGCGTTAAG CCCGACACAA TGAAGCTTGT
Idaea_virgulata_NPV402          AGTCGGGGGC CGGGGAAGAA CCAAAAACTA ACACTCTTTA AAGAAATCCG CAATGTCAAA CCCGATACGA TGAAGCTCAT
Junonia_coenia_GV19             TTCGTGGGGC CGGGGAAGAA CGTTCGCATC ACCCTCTTCA AGGAAATCAG ACGCGTCTGT CCCGATACTA TGAAACTCGT
Kotochalia_junodi_NPV205        TTCGTGGGGC CGGGGAAGAA TTTAAAACTC ACTCTGTTCA AAGAGATCCG AAACGTTAAA CCCGACACCA TGAAACTGGT
Kotochalia_junodi_NPV32         GTCGTGGGGC CGGGGAAGAA TCTAAAACTC ACTCTGTTCA AAGAGATCCG AAACGTTAAA CCCGACACCA TGAAACTGGT
Lacanobia_orecalcea_GVA6        TTCGTGGGGC CGGGGAAGAA CGTTAGAATC ACGCTGTTCA AGGAGATTCG TAGGGTTCAC CCTGATACCA TGAAGCTTAT
Lasiocampa_quercus_NPV202       TTCGGGGGGC CGGGGAAGAA CCAAAAACTT ACTCTTTTCA AGGAAATTAG AAATGTAAAG CCCGACACGA TGAAGCTCAT
LdMNPV                          TTCTACGGGC CCGGCAAAAA TCAGAAACTC ACTCTCTTCA AAGAGATCCG CAATGTCAAG CCCGACACGA TGAAGCTGGT
Lonomia_obliqua_NPV             NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNA AGAAATTCGT AACGTTAAAC CCGGACACAA TGAAACTTAT
Lymantria_monacha_NPVB4         TTCGTGGGGC CGGGGAAGAA CCATAAACTC ACCCTCTTCA AAGAAAGTAG CAATGTGAAG CCCGACACGA TGAAGCTGAT
Lymantria_nanayi_NPV750         TTGGGGGGGC CGGCGAAGAA TCAGAACCTC ACTCTCTTCA AAGAGAGCCG CAATGTCaAG CCTGACAATA TGGAGCTGGT
Lymantria_obfuscata_NPV193      TTCGTGGCGC CGGGGAAGAA TCAGAAACTC ACTCTCTTCA AAGAGATCcG CAATGTCAAG CCCGACACGA TGAAGCTGGT
Mahasena_corbetti_GV699         GTCGTGGGGC CGGGGAAGAA CCAAAAACTT CAACTCTTCA AAGAGATGCG TACTGTCAAA CCCGACACCA TGAAACTAGT
Mahasena_corbetti_NPV595        NNAGTGGGGC CGGGGAAGAA TCAAAAATTA ACTCTTTTCA AAGAGATACG AAACGTCAAA CCCGACACCA TGAAACTAGT
Mahasena_corbetti_NPV692        TTCGGGGGGC CGGGGAAGAA TCAAAACATT AACTCTTTTC ACGAGATACG AAACGTCAAA CCCGACACCA TGAAACTAGT
Malacosoma_neustria_NPV         TTCTTGGGGC CTGGCAAAAA CCAAAAACTC ACATTGTTCA AGGAAATTCG TAATGTCAAA CCGGACACGA TGAAACTCAT
Malacosoma_neustria_NPVk12      TTCGGGGGGC CGGGGAAGAA CCAAAAACTC ACATTGTTCA AAGAAATCCG TAATGTCAAA CCGGACACGA TGAAGCTCAT
Malacosoma_pluviale_NPVk3       NNNNNNNNNN NNNNNNCGAT CCAAAAACTC ACATTGTTCA AAGAAATCCG TAATGTCAAA CCGGACACGA TGAAGCTCAT
```

```
Malocosoma_disstria_NPV          TTCCTCGGAC CCGGCAAAAA TCAAAAGCTC ACACTGTTCA AAGAGATACG CAATGTCAAA CCGGACACGA TGAAGCTCAT
Mamestra_brassicae_NPV           TTCCTGGGAC CGGGCAAAAA CCAAAAACTA ACTTTGTTTA AGGAAATCAG AAATGTTAAG CCCGACACCA TGAAGCTTGT
Mamestra_brassicae_NPV274        TTCGTGGGGC CGGGGAAGAA CCAAAAACTC CCTCTCTTCA AGGAAATCCG TAATGTTAAA CCCGACACGA TGAAGCTTGT
Mamestra_brassicae_NPVK9         NNNNNNNNNN NNNNNNNNNN NNNNAAACTA ACTTTGTTTA AGGAAATCAG AAATGTTAAG CCCGACACCA TGAAGCTTGT
Mamestra_configurata_NPVA        TTCTTGGGAC CGGGCAAAAA CCAAAAGCTG ACTCTCTTCA AAGAGATCAG AAATGTTAAG CCCGACACGA TGAAGCTCGT
Mamestra_configurata_NPVB        TTCCTGGGAC CGGGCAAAAA CCAAAAACTA ACTTTGTTTA AGGAAATCAG AAATGTTAAG CCCGACACCA TGAAGCTTGT
Melanchra_persicariaGV26         TCGTGGGGGC CGGGGAAGAA CGTTAGAATC ACTCTGTTCA AGGAGATTCG TAGAGTTCAC CCCGATACCA TGAAGCTTAT
Mynthimna_pseudolatia_NPVA1      TTCGTGGGGC CGGGGAAGAA CCAAAAATTG ACTCTCTTCA AGGAGATTCG CAATGTTAAG CCCGACACCA TGAAACTCGT
'Mynthimna separata (Lese)NPV'   TTCCTGGGAC CGGGCAAAAA CCAAAAACTA ACTTTGTTTA AGGAAATCAG ATATGTTAAG CCCGACACCA TGAAGCTTGT
Mynthimna_separata_ChinaNPVA3    TTCGTGGGGC CGGGGAAGAA CCAAAAGTTG ACCCTCTTCA AAGAAATCCG TAATGTCAAA CCCGACACCA TGAAACTGAT
Mynthimna_separata_IndiaNPVA4    TTCGTGGGGC CGGGGAAGAA CCAAAAGTTG ACCCTCTTCA AAGAAATCCG TAATGTCAAG CCCGACACCA TGAAACTGAT
Mynthimna_separata_Japan_NPVA2   TTCGTGGGGC CGGGGAAGAA CCAAAAACTA ACTTTGTTTA AGGAAATCAG AAATGTTAAG CCCGACACCA TGAAGCTTGT
Natada_nararia_GV254             TTCGTGGGGC CGGGGAAGAA TGTAAGAATT ACTTTGTTCA AAGAAATTAG GAGGGTTCAT CCCGATACTA TGAAACTCGT
Nymphalis_io_NPV353              TTCGTGGGGC CGGGGAAGAA TCAAAAACTG ACATTGTTTA AAGAGATTCG TAATGTCAAA CCCGATACGA TGAAGCTCGT
Operoptera_brumata_NPV291        TTCGTGGGGC CGGGGAAGAA CCAGAAACTC ACTCTCTTCA AAGAACTGCG TACTATCAAA CCCGATACCA TGAAACTGAT
Operoptera_brumata_NPVi4         GTCGTGGGGC CGGGGAAGAA CCAGAAACTC ACTCTCTTCA AAGAACTGCG TACTATCAAA CCCGATACCA TGAAACTGAT
Operoptera_brumata_NPVi5         GTCGTGGGGC CGGGGAAGAA CCAGAAACTC ACTCTCTTCA AAGAACTGCG TACTATCAAA CCCGATACCA TGAAACTGAT
OpMNPV                           TTCTTGGGCC CCGGCAAAAA CCAAAAACTG ACCCTGTTTA AAGAAATTCG CAACGTCAAG CCCGACACGA TGAAGCTCAT
OpSNPV                           TTTTTGGGAC CTGGCAAAAA CCAAAAACTC ACTTTGTTTA AAGAAATTCG CAATGTAAAA CCCGACACTA TGAAATTAGT
Orgya_antiqua_NPVA9              TTCGTGGGGC CGGGGAAGAA CCAAAAACTC ACTTTGTTTA AGGAAATTCG CAATGTAAAA CCCGACACTA TGAAATTAGT
Orgya_leucostigma_NPVB6          TTCGTGGGGC CGGGGAAGAA TCAAAAACTT ACTTTGTTTA AAGAAATACG CAATGTCAAG CCAGACACAA TGAAACTGGT
Orgya_mixta_NPV67                TTCGTGGGGC CGGGGAAGAA CCAAAAACTC ACTCTACTCA AAGAGATCCG CAACGTTAAG CCCGACACAA TGAAACTGGT
Orgya_pseudotsugattaNPVB5        NTCGTGGGGC CGGGGAAGAA CCAAAAACTG ACTCTGTTTA AAGAAATACG CAATGTCAAG CCAGACACAA TGAAGCTGGT
Orgyia_anartoides_NPV            TCTTTGGGAC CCGGCAAAAA CCAAAAACTA ACTCTGTTCA AAGAAATTCG CAATGTGAAA CCCGACACTA TGAAATTGGT
Pachypasa_papyri_NPV436          TTCGTGGGGC CGGGGAAGAA TCAAAAATTA ACTTTATTCA AAGAAATTAG AAATGTTAAG CCCGATACTA TGAAATTAAT
Panolis_flammea_MNPV             TTCTTGGGAC CGGGCAAAAA CCAAAAACTG ACTCTCTTTA AAGAGATTAG AAATGTTAAG CCCGACACGA TGAAGCTCGT
Panolis_flammea_NPVi3            TTCGTGGGGC CGGGGAAGAA CCAAAAACTG ACTCTCTTTA AAGAGATTAG AAATGTTAAG CCCGACACGA TGAAGCTCGT
Perina_nuda_NPV                  TTTTTAGGCC CCGGCAAAAA CCAAAAACTG ACCCTGTTTA AAGAAATTCG CAACGTCAAG CCCGACACGA TGAAGCTCAT
PhopGV                           TTCAGAGGAC CGGGAAAAAA TGTTCGCATT ACTCTGTTCA AAGAGATTAG ACGAGTTCAT CCAGACACAA TGAAACTTGT
Phthorimaea_operculella_GVC3     TTCGTGGGGC CGGGGAAGAA TGTTCGCATT ACTCTGTTCA AAGAGATTAG ACGAGTTCAT CCAGACACAA TGAAACTTGT
Phthorimaea_operculella_GVC4     TTCGTGGGGC CGGGGAAGAA TGTTCGCATT ACTCTGTTCA AAGAGATTAG ACGAGTTCAT CCAGACACAA TGAAACTTGT
Pieris_brassicae_GV              TTCCGTGGTC CCGGAAAGAA TGTTAGGATT ACCCTTTTCA AAGAATGTAG ACGAGTCGAA CCCGACACAT TGAAGCTGGT
Pieris_brassicae_GV384           TTCGTGGGGC CGGGGAAGAA TGTTAGGATT ACCCTTTTCA AAGAATGTAG ACGAGTCGAA CCCGACACAT TGAAGCTGGT
Pieris_rapaeGV95                 TTCGTGGGGC CGGGGAAGAA TGTTAGGATT ACCCTTTTCA AAGAATGTAG ACGAGTCGAA CCCGACACAT TGAAGCTGGT
Plodia_interpunctella_GVB3       TTCGTGGGGC CGGGGAAGAA TGTGCGAATC ACCCTATTCA AAGAAATACG CAGAATCCAC CCCGACACCA TGAAGCTGGT
Plusia_gammaNPV287               TTCGTGGGGC CGGGGAAGAA CCAAAAACTC ACTCTCTTCA AGGAAATCCG TAATGTTAAA CCCGACACGA TGAAGCTTGT
Plusia_orichalcea_NPV            TTCTTGGGTC CCGGTAAGAA CCAGAAACTC ACTTTGTTCA AAGAAATCCG TAATGTAAAG CCTGATACCA TGAAGCTTGT
'Polygonia c-album NPV380'       TTCGTGGGGC CGGGGAAGAA TCAAAAACTG ACATTGTTTA AAGAGATTMG TAATGTCAAA CCCGATACGA TGAAGCTCGT
PxGV                             TTCCGAGGAC CCGGAAAGAA CGTGAGGATC ACATTGTTCA AGGAGATCCG CAGGGTGCAA CCCGACACTC TTAAGCTTGT
RaouNPV                          TTTCTAGGAC CTGGCAAGAA CCAAAAACTC ACTCTTTTCA AAGAAATCCG TAATGTTAAA CCCGATACCA TGAAGTTGAT
Samia_ricini_NPV28               GTTCGGGGGC CGGGGAAGAA CCAAAAACTC ACTCTCTTCA AGGAAATCCG TAATGTTAAA CCCGACACGA TGAAGCTTGT
Shpinx_ligustri_NPV334           NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NAATGTGAAA CCCGATACCA TGAAGTTAAT
Smerinthus_ocellata_NPV819       TTCGGGGGGC CGGGGAAGAA TCAAAAGTTG ACTCTGTTCA AGGAGATCTG TAACGTGAAA CCCGACACGA TGAAGCTGAT
Sphinx_ligustri_NPV362           TTCGTGGGGC CGGGGAAGAA CCAAAAACTC ACTCTCTTCA AGGAAATCCG TAATGTTAAA CCCGACACGA TGAAGCTTGT
Spilosoma_lutea_GV126            TTCGTGGGGC CGGGGAAGAA TGTTAGAATT ACCTTGTTCA AGGAAATTCG CCGAGTCCAA CCGGACACAC TGAAATTGGT
Splisoma_imparilis_NPVN2         NNNNNNNNNN NNNNNNNNAA CCAAAAATTG TCGCTTTTTA AAGAAATCCG CAACGTCAAG CCCGACACAA TGAAACTGAT
Spodoptera__exempta_NPVK11       TTCGTGGGGC CGGGGAAGAA TCAGAAGCTG ACGCTTTTCA AAGAAATCCG TAACGTCAAG CCCGACACGA TGAAGCTCGT
Spodoptera_exigua_MNPV           TTCTTGGGAC CCGGCAAAAA CCAAAAGTTG ACTCTCTTCA AGGAAATTCG TATCGTCAAA CCGGACACGA TGAAACTGGT
Spodoptera_frugiperda_NPV        TTCCTGGGAC CCGGCAAAAA CCAAAAGCTG ACTCTGTTCA AAGAGATCCG TATCGTCAAG CCCGACACCA TGAAGCTGGT
Spodoptera_frugiperda_NPVK10     TTGGGGGGGC CGGGGAAATA CCAAAAGCTA ACTCTGTTCA AAGAAATATG TATTGTTAAG CCCGACACCA TGAAGCTGGT
Spodoptera_litoralis_NPV         TTCATGGGTC CCGGTAAAAA TCAAAAGTTG ACTCTGTTCA AGGAGATCCG TAACGTGAAA CCCGACACGA TGAAGCTGAT
Spodoptera_littoralis_GV66       GTCGTGGGGC CGGGGAAGAA TGTTAAAATC ACTCTATTCA AAGAGATTCG TCGTATTCAT CCAGACACTA TGAAGCTAGT
Spodoptera_littoralis_NPV440     NNNNNNNNNN NNNNNNNNNA TCAAAAGTTG ACTCTGTTCA AGGAAATGCG TAACGTGAAA CCCGACACGA TGAAGCTGAT
Spodoptera_litura_GV766          TTCGGGGGGC CGGGGAAGAA TGTTAAAATC ACTCTATTCA AAGAGATTCG TCGTATTCAT CCAGACACTA TGAAGCTAGT
Spodoptera_litura_NPVg           TTCCTCGGAC CCGGCAAGAA CCAAAAGTTG ACTCTGTTTA AAGAAATCCG TCTCGTCAAG CCCGACACGA TGAAACTGGT
Spodoptera_litura_NPV2           TTCATGGGTC CTGGTAAAAA TCAAAAGTTG ACTCTGTTCA AAGAGATTCG TAACGTGAAG CCCGACACGA TGAAGCTGAT
Spodoptera_sunia_NPVK8           NNNNNNNNNN NNNNNNNAAA TCCTAAACTA AATTTGTTTA AGGAAATTTG TAATGTTAAG CCCGATACCA TGAAGCTCGT
Spodoptera_sunia_NPVK7           NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NGTTGTCAAA CCCGACACAA TGAAATTAGT
Thysanoplusia_orichalcea_NPV     TTCCTTGGAC CCGGCAAGAA CCAAAAATTG ACGTTATTTA AAGAAATTCG TAATGTCAAA CCCGACACCA TGAAATTGAT
Thysanoplusia orichalcea NPV2    TTCTTgGGtC CCGGtAAGAA CCAgAAAcTc ACtTTgTTcA AAGAAATcCG TAATGTaAAg CCtGAtACCA TGAAgcTtgT
Trabala_vishnou_NPV195           TTCGTGGGGC CGGGGAAGAA TCAAAAATTG ACTTTGTTTA AAGAAATCCG CAATGTCAAA CCGGACACGA TGAAACTGGT
Trichoplusia_ni_GV               TTCAGAGGCC CTGGTAAGAA TGTTAAGATT ACCCTGTTTA AGGAGATTCG TCGGATCCAG CCAGATACCA TGAAGCTTGT
Trichoplusia_ni_MNPVC2           GTTCGGGGGC CGGGGCAAGA CCAAAAACTC ACTCTCTTCA AGGAAATCCG TAATGTTAAA CCCGACACGA TGAAGCTTGT
Trichoplusia_ni_SNPV             TTCCTGGGAC CCGGTAAGAA CCAAAAACTC ACTTTGTTTA AAGAGATCCG TAATGTAAAG CCCGATACCA TGAAGCTTGT
Trichoplusia_orichalceaNPVB9     TTCGTGGGGC CGGGGAAGAA CCAAAAACTC ACTCTTTTCA AGGAAATTCG TAATGTTAAG CCCGACACCA TGAAGCTCGT
Utetheisa_pulchella_NPV142       NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNTATCAAA CCCGACACCA TGAAACTGAT
Vanessa_alatanta_NPV54           TTCGTGGGGC CGGGGAAGAA TCAAAAACTG ACATTGTTTA AAGAGATTCG TAATGTCAAA CCCGATACGA TGAAGCTCGT
Vanessa_cardui_NPV367            TTCGGGGGGC CGGCGAAGAA TCAAAAaCTG ACATtGTTTA AAgAGATTAG TAATGTCAAA CCCGATACGA TGAAGCTCGT
Wiseana_cervinata_GV342          TTCGTGGGGC CGGGGAAGAA TGTGAGAATT ACTCTTTTCA AAGAAATTCG CAGGGTGCAG CCCGATACTA TCAAACTCGT
Wiseana_cervinata_NPV344         TTCGTGGGGC CGGGGAAGAA TCAGAAACTT ACTTTGTTCA AAGAGATTAG GAACATCAAG CCCGACACGA TGAAGCTCAT
Wiseana_signata_NPV              TTCTGCGGGC CGGGTAAGAA TCAGAAACTT ACTTTGTTCA AAGAGATTAG GAACATCAAG CCCGACACGA TGAAACTCAT
XcGV                             TTCCGCGGTC CTGGTAAAAA TGTAAAAATC ACCCTGTTCA AAGAGATTCG CCGGATCCAG CCCGACACCA TGAAGCTCGT

Abraxasgrossulariata_NPV112      CGTCAACTGG AGCGGTAAAG AGTTTTTGCG CGAGACATGG ACCCGCTTCG TTGAAGACAG CTTCCCCATC GTCAACGACC
Achaea_faber_NPV64               CGTAAACTGG AGCGGCAAAG AGTTTCTCAG GGAAACTTGG ACCCGCTTTA TGGAGGACAG CTTTCCCATT GTCAACGACC
AchaeajanataGV835                GTGCAACTGG ACCGGTAAAG AGTTTCTGCG CGAATCATGG ACCCGCTTCA TTTCAGAGGA GTTTCCCATC ACCACCGACC
AcMNPV                           CGTTGGATGG AAAGGAAAAG AGTTCTACAG GGAAACTTGG ACCCGCTTCA TGGAAGACAG CTTCCCCATT GTTAACGACC
AcMNPVK1                         CGTTGGATGG AAAGGAAAAG AGTTCTACAG GGAAACTTGG ACCCGCTTCA TGGAAGACAG CTTGCCCATT GTTAACGACC
Actias_selene_NPV47              TGTCAACTGG AGCGGTAAAG AATTTCTGCG CGAAACTTGG ACCCGTTTTG TTGAGGATAG CTTTCCGATT GTAAACGACC
Adoxophyes_orana_GV              GTGCAACTGG AGCGGGAAAG AGTTTCTGAG AGAGACTTGG ACTCGTTTCA TTTCCGAAGA ATTCCCGATC ACCACCGATC
AfMNPVK2                         CGTCAACTGG AGCGGCAAAG AGTTTATGCG TGAAACTTGG ACCCGTTTCG TTGAGGACAA GTTTCCCATT GTAAATGACC
Aglais_urticae_NPVA7             CGTCAACTGG TCCGGTAAAG AATTCCTCAG GGAAACTTGG ACTCGGTTCA TGGAAGACAG TTTCCCCATC GTTAACGACC
Agraulis_vanillare_NPV779        AGTCAATTGG AGCGGCAAAG AGTTTATGCG CGAAACTTGG ACCCGTTTCG TTGARGACAG TTTTCCCATT GTYAACGACC
Agrotis_segetum_NPVi6            TGTTAACTGG AGCGGCGAAG AGTTTCTCAG GGAAACTTGG ACCCGTTTCA TGGAGGACAG CTTTCCAATT GTTAATGACC
Agrotis_segetum_NPVK4            TGTTAACTGG AGCGGGAAAG AGTTTCTCAG GGAAACTTGG ACCCGTTTCA TGGAGGACAG CTTTCCAATT GTTAATGACC
Ammalo_insulata_NPV175           TGTCAACTGG AGCGGCAAAG AGTTTCTGCG CGAAACCTGG ACCCGTTTCG TCGAAGACAG TTTTCCCATC GTAAACGATC
Amonis_sabulifera_NPV14          CGTCAACTGG AGCGGTAAAG AATTTTTGAG GGAAACTTGG ACTCGTTTCA TGGAGGACAG TTTCCCTATC GTTAACGACC
Amsacta_albistriga_NPV           CGTCAATTGG AGCGGCAAAG AGTTTCTTCG TGAGACTTGG ACTCGTTTCG TTGAAGACAG TTTCCCCATT GGTAACGATC
Antherea_pernyi_NPV              TGTCAACTGG AGCGGTAAAG AATTTCTGCG CGAAACTTGG ACCCGCTTTG TTGAGGATAG CTTTCCGATT GTAAACGACC
Antherea_polyphemus_NPV30        CGTCAACTGG AGCGGAAAAG AGTTTCTGCG TGAAACTTGG ACCCGTTTCG TTGAGGACAG TTTCCCTATT GTAAACGACC
Anthophila_fabricana_GV790       CTGCAATTGG AGCGGTAAAG AGTTTCTTCG TGAAACATGG ACGAGATTCG TATCTGAAGA ATTTCCCATC ACCACCGACC
Anticarsia_gemnitalis_NPV        TGTTAACTGG AGCGGTAAAG AGTTTCTGCG CGAAACTTGG ACTCGTTTTG TTGAAGACAG CTTCCCCATT GTAAACGACC
Aporia_crataegi_NPV106           CGTAAACTGG AGCGGCAAAG AGTTTCTGCG TGAAACTTGG ACACGTTTTG TTGAGGACAG TTTTCCAATA GTAAACGATC
Archips_cerasivora_NPV           TGTTAACTGG AGCGGGAAAG AGTTTCTGGG GCAAACCTGG ACGCGCTTCG TTGAAGACAG CTTCCCTATT GTTAACGACC
Aroa_discalis_NPV63              CGTCAACTGG AGCGGCAAAG AGTTTCTCAG GGAAACTTGG ACTCGTTTCA TGGAAGACAG TTTCCCCATT GTTAATGACC
Artogeia_napi_GV242              GTGCAACTGG AGCGGTAAAG AATTTCTTCG TGAAATGTGG ACTCGTTTTA TCTCAGAAGA GTTCCCAATC ACCACCGACC
Attacus_ricini_NPV               TGTCAACTGG AGCGGTAAAG AGTTTTTGCG CGAAACTTGG ACCCGCTTTG TTGAAGATAG CTTTCCGATT GTAAACGACC
BmNPV                            CGTCAACTGG AGCGGCAAAG AGTTTTTGCG TGAAACTTGG ACCCGTTTTG TTGAGGACAG CTTCCCCATT GTAAACGACC
Bombyx_mori_NPV460               CGTCAACTGG AGCGGCAAAG AGTTTCTGCG TGAAACTTGG ACCCGTTTTG TTGAGGACAG CTTCCCCATT GTAAACGACC
Buzura_edwardsii_NPV505          AGTAAACTGG AGCGGCAAAG AGTTTCTCAG GGAAACTTGG ACCCGTTTCA TGGAAGACAG CTTCCCTATC GTGAACGACC
Buzura_Supressaria_NPV           CGTTAACTGG AGCGGTAAAG AATTTCTGAG GGAAACTTGG ACCCGCTTTA TGGAGGATAG TTTTCCTATT GTGAACGACC
Cacoecia_murinana_GV37           ATGCAACTGG AGCGGAAAAG AATTCCTTCG CGAAACATGG ACACGCTTCA TTTCTGAAGA GTTTCCCATC ACTACTGACC
Catopsila_pomona_NPV78           CGTCAATTGG AGCGGCAAAG AGTTTTTGCG TGAAACTTGG ACCCGTTTCG TCGAAGACAG TTTTCCAATT GTTAACGACC
Cerapteryx_gramini_NPVA5         CGTCAACTGG AGCGGGAAAG AGTTYCTGAG GGAAACTTGG ACCCGTTTCA TGGAGGACAG CTTTCCCATC GTCAACGAYC
```

```
ChfuNPV                          TGTCAACTGG AGCGGTAAAG AGTTTTTCCG CGAAACCTGG ACCCGGTTCG TTGAAGACAG CTTTCCCATT GTTAACGACC
ChfuNPVDEF                       TGTTAACTGG AGCGGTAAAG AGTTTCTGCG CGAAACTTGG ACTCGTTTCG TTGAAGACAG TTTCCCCATT GTAAACGACC
Choristoneura_furamifera_GV      ATGCAACTGG AGCGGAAAGG AATTCCTTCG CGAAACATGG ACACGCTTCA TTTCTGAAGA GTTTCCCATC ACTACTGACC
Choristoneura_rosaceana_NPV      TGTCAACTGG AGCGGTAAAG AGTTTTTGCG CGAAACCTGG ACCCGGTTCG TTGAAGACAG CTTTCCCATT GTTAACGACC
Cirphis_unipuncta_NPV172         CGTCAACTGG AGCGGCAAAG AGTTTCTCAG GGAAACTTGG ACCCGTTTCA TGGAAGACAG CTTTCCGATC GTTAACGACC
CpGV                             GTGCAACTGG AGCGGCAAAG AATTCCTTCG TGAAACCTGG ACCCGCTTCA TTTCCGAAGA GTTCCCCATC ACCACCGACC
Cryptophlebia_leucotreta_GV      ATGCAACTGG AGCGGCAAAG AATTCCTTCG CGAGACTTGG ACCCGTTTCA TCTCTGAAGA ATTTCCCATC ACCACAGACC
Cynosarga_chrysolopa_NPV302      TGTCAACTGG AGCGGTAAAG AGTTTTTGCG TGAAACTTGG ACCCGCTTTG TTGAAGACAG TTTTCCCATT GTTAACGACC
Darna_trima_GV545                GTGCAATTGG AGCGGTAAAG AATTTTTGCG TGAAACTTGG ACTAGATTCA TTTCCGAAGA GTTTCCAATA ACTACCGATC
Diacrisia_obliqua_NPV798         CGTCAACTGG AGCGGCAAAG AGTTTCTGCG TGAAACCTGG ACCCGCTTTG TTGAAGACAG TTTTCCCATT GTAAACGACC
Diatrea_saccharalis_NPVK6        GTGTAATTGG AGCGGAAAAG AATTTCTACG CGAAACTTGG ACTAGATTTA TTTCTGAGGA GTTTCCAATT ACGACCGATC
Ectropis_obliqua_NPV             CGTTAATTGG AGCGGCAAAG AATTTCTCAG AGAAACTTGG ACTCGTTTCA TGGAAGACAG TTTCCCCATT GTCAACGACC
Ephestia_cautella_NPVC5          TGTCAACTGG AGTGGTAAAG AGTTTCTGCG TGAAACGTGG ACTCGTTTCG TTGAAGACAG TTTCCCCATT GTAAATGATC
EppoMNPV                         TGTCAACTGG AGCGGCAAAG AGTTTCTGCG TGAAACTTGG ACTCGATTCG TTGAGGATAG CTTCCCCATT GTAAACGATC
Erannis_defolariaNPV174          TGTCAACTGG AGCGGAAAAG AGTTTCTTAG GGAAACTTGG ACTAGGTTTA TGGAGGACAG TTTCCCTATT GTGAACGATC
Euplasia_luciparaGV248           CTGCAACTGG AGCGGTAAGG AATTTTTGCG TGAGGCCTGG ACCAGATTCA TCTCCGAGGA GTTCCCCATC ACCACTGACC
Euproctis_Chrysorrohea_NPVA10    CGTAAATTGG AGCGGCAAAG AGTTTCTGCG TGAAACTTGG ACTCGTTTTG TAGAAGATAG TTTTCCCATT GTCAACGACC
Euproctis_similis_NPV768         CGTAAATTGG AGCGGCAAAG AGTTTCTGCG TGAAACTTGG ACTCGTTTTG TAGAAGATAG TTTTCCCATT GTCAACGACC
Harrisina_brillians_GV           GTGCAATTGG AGCGGAAAGG AATTTTTGCG CGAAACATGG ACGCGATTTA TTTCTGAGGA GTTTCCAATC ACTACCGATC
Harrisina_brillians_GVM2         GTGCAATTGG AGCGGAAAGG AATTTTTGCG CGAAACATGG ACGCGATTTA TTTCTGAGGA GTTTCCAATC ACTACCGATC
HaSNPV                           AGTTAACTGG AGCGGTCGCG AATTTCTTCG CGAAACTTGG ACGCGTTTCA TGGAAGACAG TTTTCCCATT GTAAACGACC
Heliconius_erato_NPV789          AGTCAACTGG AGCGGCAAAG AGTTTATGCG CGAAACTTGG ACGCGTTTCG TTGAGGACAG TTTTCCCATT GTCAACGACC
Homona_coffearia_GV744           ATGCAACTGG AGCGGAAAGG AATTCCTTCG CGAAACATGG ACACAATTCA TTTCTGAAGA GTTTCCCATT ACTACCGACC
Homona_coffearia_NPV745          ATGCAACTGG AGCGGAAAGG AATTCCTTCG CGAAACATGG ACACGCTTCA TTTCTGAAGA GTTTCCCATT ACTACCGACC
Hyphantria_cunea_GV228           TTGCAATTGG AGCGGAAAGG AATTTTTACG CGAAACTTGG ACTCGGTTTA TTTGCGAGGA GTTCCCCATC ACTACAGACC
Hyphantria_cunea_NPV             CGCTAACTGG AGCGGCAAAG AGTTTCTGCG TGAAACCTGG ACCCGCTTTG TTGAAGACAG TTTTCCCATT GTAAACGACC
Hyphantria_cunea_NPVN3           CGTCAACTGG AGCGGCAAAG AGTTTCTGCG TGAAACCTGG ACCCGCTTTG TTGAAGACAG TTTTCCCATT GTAAACGACC
HzSNPV_genome                    AGTTAACTGG AGCGGTCGCG AATTTCTTCG CGAAACTTGG ACTCGTTTCA TGGAAGACAG TTTTCCCATT GTAAACGACC
Idaea_virgulata_NPV402           TGTCAATTGG AGCGGTAAAG AGTTTCTGCG CGAAACTTGG ACCCGCTTCG TTGAAGACAG CTTCCCCATT GTTAACGATC
Junonia_coenia_GV19              CTGCAATTGG AGTGGTAAAG AATTTCTACG CGAAACTTGG ACACGATTCA TCAGCGAAGA ATTTCCCATC ACCACCGACC
Kotochalia_junodi_NPV205         GGTAAATTGG AGTGGTAGGG AATTCCTCCG CGAAACATGG ACCCGATTTA TGGAAGACAG TTTCCCTATT GTAAACGACG
Kotochalia_junodi_NPV32          GGTAAATTGG AGTGGTAGGG AATTCCTCCG CGAAACATGG ACCCGATTTA TGGAAGACAG TTTCCCTATT GTAAACGACG
Lacanobia_orecalcea_GVA6         TTGCAATTGG AGCGGCAAAG AATTCCTTAG GGAAACCTGG ACCCGTTTCA TCTCTGAAGA GTTCCCCATC ACTACCGACC
Lasiocampa_quercus_NPV202        CGTCAATTGG AGCGGCAAAG AGTTTTTACG TGAAACTTGG ACGCGTTTCA TGGAGGACAG CTTTCCTATT GTCAACGATC
LdMNPV                           GGTCAACTGG AGCGGCAAAG AGTTTCTCAG GGAGACTTGG ACCCGCTTCA TGGAGGACAG CTTCCCCATC GTCAACGACC
Lonomia_obliqua_NPV              TGTCAACTGG AGCGGAAAAG AGTTTCTTCG CGAAACTTGG ACCCGTTTTG TTGAGGACAG CTTCCCAATT GTAAACGACC
Lymantria_monacha_NPVB4          GGTCAACTGG AGCGGTAAAG AGTTTCTGCG CGAAACTTGG ACCCGCTTTA TGCAGGACAG GTTCCCCATC GTGAACGACC
Lymantria_nanayi_NPV750          GGTCAACTGG AGCGGCAAAG AGTTTCTCAG GGAGACTTGG ACCCGCTTCA TGGAGGACAG CTTTCCCATC GTCAACGACC
Lymantria_obfuscata_NPV193       GGTCAACTGG AGCGGCAAAG AGTTTCTCAG GGAGACTTGG ACCCGCTTCA TGGAGGACAG CTTCCCCATC GTCAACGACC
Mahasena_corbetti_GV699          TGTTAATTGG AGCGGTAAAG AGTTCCTCAG GGAGACTTGG ACTCGTTTCA TGGAGGACAG CTTCCCCATT GCCAACGATC
Mahasena_corbetti_NPV595         TGTTAATTGG AGCGGTAAAG AGTTTCTAAG GGAGACTTGG ACACGTTTCA TGGAAGACAG CTTCCCGATA GTTAACGATC
Mahasena_corbetti_NPV692         TGTTAATTGG AGCGGTAAAG AGTTTCTAAG GGAGACTTGG ACACGTTTCA TGGAAGACAG CTTCCCGATA GTTAACGATC
Malacosoma_neustria_NPV          CGTAAACTGG AGCGGCAAAG AGTTTCTGCG TGAAACTTGG ACCCGTTTCA TGGAAGACAG CTTCCCCATT GTCAACGACC
Malacosoma_neustria_NPVk12       TGTGAACTGG AGCGGCAAAG AGTTTTTGCG TGAGACTTGG ACCCGATTCA TGGAAGACAG CTTTCCCATT GTCAACGACC
Malacosoma_pluviale_NPVk3        TGCGAACTGG GAGCGTAAGG AGTTTTTGCG TGAGACTTGG ACCCGATTCA TGGAAGACAG TTTCCCCATT GTCAACGACC
Malocosoma_disstria_NPV          CGTGAACTGG AGCGGTAAAG AGTTTTTGCG TGAAACTTGG ACACGTTTCA TGGAAGACAG CTTCCCCATT GTCAACGACC
Mamestra_brassicae_NPV           CGTAAACTGG AGCGGTAAAG AGTTTCTCAG AGAAACTTGG ACCCGTTTCA TGGAAGACAG CTTCCCCATT GTTAACGACC
Mamestra_brassicae_NPV274        CGTTGGATGG AACGGCAAAG AGTTTTTCAG GGAAACTTGG ACCCGTTTCA TGGAAGACAG CTTTCCCATT GTTAACGACC
Mamestra_brassicae_NPVK9         CGTAAACTGG AGCGGGAAAG AGTTTCTCAG AGAAACTTGG ACCCGTTTCA TGGAAGACAG CTTCCCCATT GTTAACGACC
Mamestra_configurata_NPVA        CGTGAACTGG AGCGGCAAAG AGTTTCTCAG GGAAACTTGG ACCCGTTTCA TGGAAGACAG CTTTCCCATC GTTAACGACC
Mamestra_configurata_NPVB        CGTAAACTGG AGCGGTAAAG AGTTTCTCAG AGAAACTTGG ACCCGTTTCA TGGAAGACAG CTTCCCCATT GTTAACGACC
Melanchra_persicariaGV26         TTGCAACTGG AGCGGCAAAG AGTTTCTTAG GGAAACTTGG ACCCGTTTCA TCTCAGAAGA GTTTCCAATC ACTACAGACC
Mynthimna_pseudolatia_NPVA1      CGTCAACTGG AGCGGCAAAG AGTTTCTGAG GGAAACTTGG ACGCGCTTCA TGGAAGACAG CTTCCCGATC GTCAACGATC
'Mynthimna separata (Lese)NPV'   CGTAAACTGG AGCGGTAAAG AGTTTCTCAG GGAAACTTGG ACCCGTTTCA TGGAAGACAG CTTCCCCATT GTTAACGACC
Mynthimna_separata_ChinaNPVA3    CGTCAACTGG AACGGCAGGG AGTTTCTGCG CGAAACTTGG ACCCGTTTCA TGGAGGACAG CTTCCCCATC GTCAACGATC
Mynthimna_separata_IndiaNPVA4    CGTCAACTGG AACGGCAGGG AGTTTCTGCG CGAAACTTGG ACCCGTTTCA TGGAGGACAG CTTCCCCATC GTCAACGATC
Mynthimna_separata_Japan_NPVA2   CGTAAACTGG AGCGGTAAAG AGTTTCTCAG GGAAACTTGG ACCCGTTTCA TGGAAGACAG CTTCCCCATT GTTAACGACC
Natada_nararia_GV254             ATGTAATTGG AGCGGTAAAG AATTTTTGCG CGAAACATGG ACTCGTTTTA TTTCAGAAGA ATTTCCAATA ACCACAGACC
Nymphalis_io_NPV353              CGTCAACTGG TCCGGTAAAG AATTCCTCAG GGAAACTTGG ACTCGGTTCA TGGAAGACAG TTTCCCCATC GTTAACGACC
Operoptera_brumata_NPV291        CGTCAACTGG TCCGGTAAAG AGTTCCTCAG GGAAACTTGG ACTCGTTTCG CAGAGGACAG CTTCCCCATT GCCAACGACC
Operoptera_brumata_NPVi4         GGTCAACTGG TCCGGTAAAG AGTTCCTCAG GGAGACTTGG ACTCGTTTCG CAGAGGACAG CTTCCCCATT GCCAACGACC
Operoptera_brumata_NPVi5         CGTCAACTGG TCCGGTAAAG AGTTCCTCAG GGAAACTTGG ACTCGTTTCG CAGAGGACAG CTTCCCCATT GCCAACGACC
OpMNPV                           CGTCAACTGG AGCGGCAAGG AGTTTTTACG TGAAACCTGG ACGCGCTTCG TGGAAGACAG CTTCCCCATC GTCAACGACC
OpSNPV                           AGTAAACTGG AGCGGCAAAG AGTTTCTGAG GGAAACTTGG ACGCGCTTCA TGGAGGACAG TTTTCCTATT GTTAACGATC
Orgya_antiqua_NPVA9              AGTAAACTGG AGCGGCAAAG AGTTTCTGAG GGAAACTTGG ACGCGCTTCA TGGAGGACAG TTTTCCTATT GTTAACGATC
Orgya_leucostigma_NPVB6          AGTGAACTGG AGCGGAAAAG AGTTTTTGAG GGAAACTTGG ACGCGATTTA TGGAAGACAG TTTCCCAATC GTGAACGATC
Orgya_mixta_NPV67                CGTCAACTGG AGCGGCAAAG AGTTTCTCAG GGAAACTTGG ACTCGTTTCA TGGAAGACAG TTTCCCCATT GTTAATGACC
Orgya_pseudotsugattaNPVB5        AGTGAACTGG AGCGGAAAAG AGTTTTTAAG GGAAACTTGG ACGCGATTTA TGGAAGACAG TTTCCCCATC GTGAACGACC
Orgyia_anartoides_NPV            AGTTAATTGG AGCGGTAAAG AGTTTCTTAG GGAAACTTGG ACTCGTTTCA TGGAAGACAG CTTCCCAATT GTAAACGATC
Pachypasa_papyri_NPV436          TGTTAATTGG AGCGGCAAAG AGTTTTTGAG AGAAACTTGG ACTCGTTTTA TGGAAGACAG TTTTCCCATT GTTAATGATC
Panolis_flammea_MNPV             CGTAAACTGG AGCGGCAAAG AGTTTCTTAG GGAAACTTGG ACCCGTTTTA TGGAAGACAG CTTTCCCATC GTTAACGACC
Panolis_flammea_NPVi3            CGTGAACTGG AGCGGCAAAG AGTTTCTCAG GGAAACTTGG ACCCGTTTCA TGGAAGACAG CTTTCCCATC GTTAACGACC
Perina_nuda_NPV                  CGTCAACTGG AGCGGCAAAG AGTTTTTGCG TGAAACCTGG ACGCGCTTCG TGGAAGACAG CTTCCCCATC GTCAACGACC
PhopGV                           GTGTAACTGG AGCGGAAAAG AATTCCTTCG TGAGACCTGG ACTCGCTTTA TCTCTGAAGA ATTCCCCATC ACCACCGATC
Phthorimaea_operculella_GVC3     GTGTAACTGG AGCGGAAAAG AATTCCTTCG TGAGACCTGG ACTCGCTTTA TCTCTGAAGA ATTCCCCATC ACCACCGATC
Phthorimaea_operculella_GVC4     GTGTAACTGG AGCGGAAAAG AATTCCTTCG TGAGACCTGG ACTCGCTTTA TCTCTGAAGA ATTCCCCATC ACCACCGATC
Pieris_brassicae_GV              GTGCAACTGG AGCGGTAAAG AATTTCTTCG TGAAATGTGG ACTCGTTTTA TCTCAGAAGA GTTCCCAATC ACCACCGACC
Pieris_brassicae_GV384           GTGCAACTGG AGCGGTAAAG AATTTCTTCG TGAAATGTGG ACTCGTTTTA TCTCAGAAGA GTTCCCAATC ACCACCGACC
Pieris_rapaeGV95                 GTGCAACTGG AGCGGTAAAG AATTTCTTCG TGAAATGTGG ACTCGTTTTA TCTCAGAAGA GTTCCCAATC ACCACCGACC
Plodia_interpunctella_GVB3       GTGCAACTGG AGCGGAAAGG AGTTCCTTCG TGAAACGTGG ACCCGGTTTA TCTCGGAGGA ATTTCCCATC ACCACAGATC
Plusia_gammaNPV287               CGTTGGATGG AAAGGAAAAG AGTTCTACAG GGAAACTTGG ACCCGCTTCA TGGAAGACAG CTTCCCCATT GTTAACGACC
Plusia_orichalcea_NPV            CGTTAACTGG AGCGGTAAAG AGTTTCTGAG GGAAACTTGG ACCCGTTTCA TGGAAGACAG CTTTCCTATT GTTAACGACC
'Polygonia c-album NPV380'       CGTCAACTGG TCCGGTAAAG AATTCCTCAG GGAAACTTGG ACTCGGTTCA TGGAAGACAG TTTCCCCATC GTTAACGACC
PxGV                             CTGCAACTGG AGCGGCAAAG AGTTTCTGAG GGAAACGTGG ACTCGTTTCA TTTCGGAAGA GTTCCCGATC ACCACTGACC
RaouNPV                          CGTCAACTGG AGCGGCAAAG AGTTTATGCG TGAAACTTGG ACCCGTTTCG TTGAGGACAG TTTCCCCATT GTAAATGACC
Samia_ricini_NPV28               CGTTGGATGG AAAGGAAAAG AGTTCTACAG GGAAACTTGG ACCCGCTTCA TGGAAGACAG CTTCCCCATT GTTAACGACC
Shpinx_ligustri_NPV334           CGTCAACTGG AGCGGAAAAG AGTTTCTGCG TGAAACTTGG ACCCGTTTCG TTGAGGACAG TTTCCCTATT GTAAACGACC
Smerinthus_ocellata_NPV819       CGTCAACTGG AACGGCAAAG AGTTTCTCCG TGAGACTTGG ACCCGTTTCA TGGAAGACAG CTTCCCCATC GTGAACGATC
Sphinx_ligustri_NPV362           CGTTGGATGG AAAGGAAAAG AGTTCTACAG GGAAACTTGG ACCCGCTTCA TGGAAGACAG CTTCCCCATT GTTAACGACC
Spilosoma_lutea_GV126            GTGCAACTGG AGCGGAAAAG AATTTTTGCG TGAAACCTGG ACACGATTTG TGTGCGAAGA GTTCCCCATC ACCACCGACC
Splisoma_imparilis_NPVN2         CGTCAACTGG AGCGGCAAAG AGTTTCTGCG TGAAACCTGG ACCCGCTTTG TTGAAGACAG TTTTCCCATT GTAAACGACC
Spodoptera__exempta_NPVK11       CGTAAACTGG AGCGGCAAAG AGTTTCTCAG GGAAACTTGG ACCCGCTTCA TGGAGGACAG CTTTCCCATT GTCAACGACC
Spodoptera_exigua_MNPV           CGTCAACTGG AGCGGCAAAG AGTTTCTCCG CGAAACTTGG ACGCGTTTCA TGGAAGACAG CTTTCCCATC GTCAACGATC
Spodoptera_frugiperda_NPV        AGTAAACTGG AGCGGCAAAG AGTTTCTCAG GGAAACTTGG ACCCGTTTCA TGGAAGACAG CTTCCCTATC GTGAACGACC
Spodoptera_frugiperda_NPVK10     AGTAAACTGG AGCGGCAAAG AGTTTCTCAG GGAAACTTGG ACCCGTTTCA TGGAAGACAG CTTCCCTATC GTGAACGACC
Spodoptera_litoralis_NPV         CGTCAACTGG AACGGCAAAG AGTTTCTCCG TGAGACTTGG ACCCGTTTCA TGGAAGACAG CTTCCCCATC GTGAACGATC
Spodoptera_littoralis_GV66       GTGCAACTGG AGCGGTAAAG AGTTTCTCCG TGAAACTTGG ACTCGTTTCA TCAGCGAAGA GTTTCCCATC ACCACTGACC
Spodoptera_littoralis_NPV440     CGTCAACTGG AACGGCGAAG AGTTTCTCCG TGAGACTTGG ACCCGTTTCA TGGAAGACAG CTTGCCCATC GTGAACGATC
Spodoptera_litura_GV766          GTGCAACTGG AGCGGTAAAG AGTTTCTCCG TGAAACTTGG ACTCGCTTCA TCAGCGAAGA GTTTCCCATC ACCACTGACC
Spodoptera_litura_NPVg           CGTCAACTGG AGCGGCAAAG AGTTTCTCAG GGAAACTTGG ACCCGTTTCA TGGAAGACAG CTTCCCATCG GTAAACGATC
Spodoptera_litura_NPV2           CGTCAATTGG AACGGTAAAG AGTTTCTGCG CGAGACTTGG ACTCGCTTTA TGGAAGACAG CTTCCCCATC GTGAACGATC
Spodoptera_sunia_NPVK8           CGTTAACTGG GGCGGGAAAA AGTTTCTTAG GGAAACTTGG ACTCTGTTCA TGGAGGACAG CTTTCCCATC GTAAACGACC
Spodoptera_sunia_NPVK7           GGTCAACTGG AGCGGAAAAG AGTTTCTTAG GGAAACTTGG ACACTTTTCA TGGAAGACAG CTTTCCTATT GTCAACGATC
Thysanoplusia_orichalcea_NPV     CGTCAACTGG AGCGGCAAGG AGTTTCTTCC TGAAACTTGG ACGCGTTTTG TTGAAGACAG CTTCCCTATT GTGAACGACC
Thysanoplusia_orichalcea_NPV2    CGTtAACTGG AGCGGtAAaG AGTTTCTgag gGAAACTTGG ACcCGTTTca TgGAAGACAG CTTtCCTATT GTtAACGACC
```

```
Trabala_vishnou_NPV195          GGTCAATTGG AGTGGTAAAG AGTTTCTCAG GGAGACTTGG ACCCGTTTCA TGGAGGATAG CTTCCCAATA GTCAACGACC
Trichoplusia_ni_GV              GTGTAACTGG AGCGGCAAAG AGTTTTTGCG CGAAACCTGG ACTCGTTTCA TCTCAGAAGA GTTCCCCATC ACCACCGATC
Trichoplusia_ni_MNPVC2          CGTTGGATGG AAAGGAAAAG AGTTCTACAG GGAAACTTGG ACCCGCTTCA TGGAAGACAG CTTCCCCATT GTTAACGACC
Trichoplusia_ni_SNPV            CGTTAACTGG AGCGGCAAAG AGTTTCTCAG GGAAACTTGG ACCCGCTTCA TGGAGGACAG CTTCCCCATC GTTAACGACC
Trichoplusia_orichalceaNPVB9    CGTCAACTGG AGCGGCAAAG AGTTTCTGAG GGAAACTTGG ACTCGTTTCA TGGAGGACAG CTTCCCCATC GTTAACGACC
Utetheisa_pulchella_NPV142      TGTCAACTGG ACCGGTAAAG AGTTCCTCAC CGAAACTTGG ACTCGTTTCA CAGAAGACAA CTTCCCCATT GCCAACGACC
Vanessa_alatanta_NPV54          CGTCAACTGG TCCGGTAAAG AATTCCTCAG GGAAACTTGG ACTCGGTTCA TGGAAGACAG TTTCCCCATC GTTAACGACC
Vanessa_cardui_NPV367           GGTCAACTGG GCCGGTAAAG AATTCCTCAG GGAAACTTGG ACTCGGTTCA TGGAAGACAG TTTCCCCATC GTTAACGACC
Wiseana_cervinata_GV342         ATGCAACTGG AGCGGTAAAG AATTTCTTCG TGAAACATGG ACGCGTTTCA TTgAAGACAA ATTTCCCATT ACCACAGATC
Wiseana_cervinata_NPV344        CGTGAATTGG TCGGGCAAAG AATTCTTGCG CGAGACTTGG ACTCGCTTCA TGGAGGACAG CTTTCCCATC GTCAACGACC
Wiseana_signata_NPV             CGTGAATTGG TCGGGCAAAG AATTCTTGCG CGAGACTTGG ACTCGCTTCA TGGAGGACAG CTTTCCCATC GTCAACGACC
XcGV                            GTGCAACTGG AGCGGCAAAG AGTTTTTGCG CGAGACCTGG ACTCGTTTCA TCTCCGAAGA GTTCCCGATC ACCACCGATC

Abraxasgrossulariata_NPV112     AGGAAGTGAT GGACGTGTTT CTAGTCGTCA ACATGCGT-- -CCCACGCGC CCCAACCGTT GCTACAAGTT TTTGGCGCAG
Achaea_faber_NPV64              AAGAGATCAT GGACGTTTTC CTAGTTGTAA ATATGCGT-- -CCCACTAAA CCTAACCGTT GTTACAGATT CCTCGCACAA
AchaeajanataGV835               AAGAAATCAT GGACTTGTGG TTTGAACTCC AACTGCGT-- -CCCATGCAA CCCAACAGGT GTTACCGATT CACCCTTCAG
AcMNPV                          AAGAAGTGAT GGATGTTTTC CTTGTTGTCA ACATGCGT-- -CCCACTAGA CCCAACCGTT GTTACAAATT CCTGGCCCAA
AcMNPVK1                        AAGAAGTGAT GGATGTTTTC CTTGCTGTCA ACATGCGT-- -CCCACTAGA CCCAACCGTT GTTACAAATT CCTGGCCCAG
Actias_selene_NPV47             AAGAGGTGAT GGATGTGTTC CTCGTCATTA ACCTGCGC-- -CCCACGCGC CCCAACAGGT GCTACAAGTT CCTGGCGCAG
Adoxophyes_orana_GV             AGGAGATCAT GGACCTGTGG TTCGAGATTC AGCTGCGA-- -CCGATGCAT CCCAATAGGT GTTACAAGTT CACGATGCAA
AfMNPVK2                        AAGAGGTGAT GGACGTGTTC CTCGCCTTCT ACCTCAGA-- -CCCACACGC CCCAATAGGT GCTTCCCGTT CCTCGATGAA
Aglais_urticae_NPVA7            AAGAAATAAT GGATGTATTC CTTGTGATCA ATATGCGT-- -CCCACTAAA CCTAATAGAT GCTACAAGTT CCTTGCTCAA
Agraulis_vanillare_NPV779       AAGAGGTAAT GGACGTGTTC CTCGTTGTCA ACCTGCGM-- -CCCACTCGT CCCAACCGAT GCTACAAATT CCTAGCCCAA
Agrotis_segetum_NPVi6           AGGAGATCAT GGACGTTTTT CTAGTTGTCA ACATGAGA-- -CCCACCAAA CCTAACCGTT GTTACAGATT TTTAGCTCAA
Agrotis_segetum_NPVK4           AGGAGATCAT GGACGTTTTT CTAGTTTTCA ACATGAGA-- -CCCACCAAA CCTAACCGTT GTTACAGATT TTTAGCTCAA
Ammalo_insulata_NPV175          AGGAGGTGAT GGACGTGTTC CTCGTTGTTA ACCTGCGC-- -CCCACGCGT CCCAACAGGT GCTACAAGTT CTTGGCACAG
Amonis_sabulifera_NPV14         AAGAAATTAT GGACGTATTC CTCGTGGTAA ACATGAGA-- -CCCACTCGT CCCAACCGTT GTTTCAAGTT TGTCGCTCAA
Amsacta_albistriga_NPV          AAGAAGTAAT GGATGTCTTT TTGGTAGTCA ACCTGCGT-- -CCCACTAGA CCCAATAGAT GTTACAAATT CCTCGCTCAG
Antherea_pernyi_NPV             AAGAGGTCAT GGATGTGTTC CTCGTCATTA ACCTGCGC-- -CCCACGCGC CCCAACAGGT GCTACAAGTT CTTGGCGCAG
Antherea_polyphemus_NPV30       AAGAGGTGAT GGACGTGTTC CTCGTCGTCA ACCTCAGA-- -CCCACACGC CCCAACAGGT GCTACAAGTT CCTCGCTCAA
Anthophila_fabricana_GV790      AAGAGATCAT GGACTTGTGG TTTGAACTTC AACTGAGA-- -CCGATGCAC CCCAATCGTT GCTACAAGTT TACCATGCAG
Anticarsia_gemnitalis_NPV       AAGAAGTGAT GGATGTGTTT TTAGTCATTA ACCTCCGC-- -CCAACTCGC CCTAACCGTT GCTACAAATT CCTGGCGCAA
Aporia_crataegi_NPV106          AAGAGGTTAT GGACGTGTTT TTAGTAATTA ACATGCGA-- -CCAACTCGT CCCAACAGAT GTTACAAATT TTTGGCGCAG
Archips_cerasivora_NPV          AAGAGGTGAT GGACGTGTTC CTCGTCATCA ACCTGCGG-- -CCCACACGG CCCAACCGCT GCTATAAGTT CCTC---CAG
Aroa_discalis_NPV63             AGGAAATTAT GGATGTGTTT CTGGTGATCA ACATGCGT-- -CCCACACGA CCAAACAGAT GCTACAGATT TTTGGCGCAA
Artogeia_napi_GV242             AACAAATAAT GGACTTGTGG TTCGAGATTC AATGTAGA-- -CCAATGCAA CCTAACCGTT GCTACAAGTT TACTATGCAA
Attacus_ricini_NPV              AAGAGGTCAT GGATGTGTTC CTCGTCATTA ACCTGCGC-- -CCCACGCGC CCCAACAGGT GCTACAAGTT CCTGGCGCAG
BmNPV                           AAGAGGTGAT GGACGTGTAC CTCGTCGCCA ACCTCAAA-- -CCCACACGC CCCAACAGGT GCTACAAGTT CCTCGCTCAA
Bombyx_mori_NPV460              AAGAGGTGAT GGACGTGTAC CTCGTCGCCA ACCTCAAA-- -CCCACACGC CCCAACAGGT GCTACAAGTT CCTCGCTCAA
Buzura_edwardsii_NPV505         AAGAAATTAT GGACGTGTTC CTAGCGATCA ACATGAGA-- -CCCACTAGG CCCAACAGAT GTTTCAGATT CCTGGCGCAA
Buzura_Supressaria_NPV          AAGAAATCAT GGATGTCTTT TTGGTAATTA ACATGCGC-- -CCAACCAGA CCTAATCGTT GTTACAGATT TTTAGCTCAA
Cacoecia_murinana_GV37          AGGAAATCAT GGATTTGTGG TTTGAAATTC AACTCAGA-- -CCAATGCAA CCCAATCGCT GTTACAAATT TACCATGCAA
Catopsila_pomona_NPV78          AAGAGGTTAT GGACGTGTTC CTCGTTCTCA ATTTGAGA-- -CCCACACGC CCCAACAGAT GTTACAAATT TTTAGCGCAA
Cerapteryx_gramini_NPVA5        AAGAAATCAT GGATGTGTTT CTGGTGGTTA ACATGAGA-- -CCCACTAGA CCCAACCGTT GTTACAGGTT CCTGGCCCAG
ChfuNPV                         AAGAAGTCAT GGACGTGTTC CTAGTCGTAA ATATGCGT-- -CCCACCCGT CCCAACCGCT GCTACAAATT CTTGGCGCAA
ChfuNPVDEF                      AGGAGGTGAT GGACGTGTTT TTAGTCATTA ACCTGCGC-- -CCAACTCGC CCTAACCGTT GCTACAAATT CCTGGCGCAA
Choristoneura_furamifera_GV     AAGAAATCAT GGATTTGTGG TTTGAAATTC AACTCAGA-- -CCAATGCAA CCCAATCGCT GTTACAAATT TACGATGCAA
Choristoneura_rosaceana_NPV     AAGAAGTCAT GGACGTGTTC CTAGTCGTAA ATATGCGT-- -CCCACGCGT CCCAACCGCT GCTACAAATT CTTGGCGCAG
Cirphis_unipuncta_NPV172        AAGAAGTTAT GGATGTGTTC CTGGTAATCA ACATGCGT-- -CCCACTAGA CCCAACCGTT GTTACAAATT CCTGGCGCAG
CpGV                            AAGAGATCAT GGACCTGTGG TTCGAGCTTC AACTGCGA-- -CCGATGCAC CCCAACCGTT GCTACAAGTT CACCATGCAG
Cryptophlebia_leucotreta_GV     AAGAAATTAT AGATTTGTGG TTTGAGCTTC AGCTACGA-- -CCAATGCAC CCTAACCGTT GTTACAAATT CACTATGCAG
Cynosarga_chrysolopa_NPV302     AAGAGGTGAT GGACGTGTTT CTCGTCGTCA ATCTGAGA-- -CCCACGCGC CCCAACAGGT GCTACAAGTT CCTTGCTCAA
Darna_trima_GV545               AAGAAATTAT GGATCTGTGG TTTGAAATTC AACTTAGA-- -CCTATGCAA CCAAACAGAT GTTACAAATT TACACTCCAA
Diacrisia_obliqua_NPV798        AAGAAGTGAT GGACGTGTTC CTGGTCATTA ACCTGCGT-- -CCCACTCGC CCCAATCGTT GCTACAAATT CTTGGCGCAA
Diatrea_saccharalis_NPVK6       AAGAAATTAT GGATTTGTGG TTCGAGCTTC AATTGAGA-- -CCAATGCAC CCTAATCGTT GTTATAAATT TACCATGCAA
Ectropis_obliqua_NPV            AAGAAGTGAT GGACGTGTTT TTGGTGATTA ATATGCGT-- -CCCACCAGG CCCAACCGTT GTTACAAATT CTTAGCTCAA
Ephestia_cautella_NPVC5         AGGAAGTTAT GGACGTTTTC CTAGTCGTCA ACTTGAGA-- -CCCACCCGT CCCAACAGAT GCTACAAATT CCTCGCTCAA
EppoMNPV                        AAGAGGTAAT GGACGTGTTT CTTGTCATTA ATTTGCGT-- -CCTACGCGC CCCAACCGTT GCTACAAGTT CTTGGCGCAA
Erannis_defolariaNPV174         AAGAAATCAT GGATGTTTTC TTGGTTGTCA ACATGCGT-- -CCCACTAGA CCCAACCGTT GCTTTAGATT TTTGGCTCAA
Euplasia_luciparaGV248          AGGAAATCAT GGACTTGTGG TTTGAGCTTC AGCTCCGA-- -CCCATGCAA CCCAACCGTT GCTACAGATT CACCATGCAG
Euproctis_Chrysorrohea_NPVA10   AAGAAGTTAT GGACGTGTTC CTGGTCATTA ATTTGAGA-- -CCGTGCAAA CCCAACAGAT GTTACAAGTT TTTAGCTCAA
Euproctis_similis_NPV768        AAGAAGTTAT GGACGTGTTC CTGGTCATTA ATTTGAGA-- -CCGTGCAAA CCCAACAGAT GTTACAAGTT TTTAGCTCAA
Harrisina_brillians_GV          AAGAAATAAT GGATTTATGG TTTGAACTTC AACTCAGA-- -CCAATGCAC CCAAATCGTT GTTATAAATT TACAATGCAA
Harrisina_brillians_GVM2        AAGAAATAAT GGATTTATGG TTTGAACTTC AACTCAGA-- -CCAATGCAC CCAAATCGTT GTTATAAATT TACAATGCAA
HaSNPV                          AAGAAATTAT GGACGTGTTT CTGTCTGTTA ATATGCGA-- -CCAACCAAA CCGAACCGTT GTTACCGATT CTTAGCGCAA
Heliconius_erato_NPV789         AAGAGGTAAT GGACGTGTTC CTCGTTGTCA ACCTGCGC-- -CCCACTCGT CCCAACCGCT GCTACAAATT CCTCGCCCAG
Homona_coffearia_GV744          AAGAAATCAT GGATTTGTGG TTTGAAATTC AACTGAGA-- -CCAATGCAA CCCAACCGCT GTTACAAATT TACCATGCAA
Homona_coffearia_NPV745         AAGAAATCAT GGATTTGTGG TTTGAAATTC AACTGAGA-- -CCAATGCAA CCCAACCGCT GTTACAAATT TACCATGCAA
Hyphantria_cunea_GV228          AGGAAATCAT GGATTTGTGG TTTGAAATTA AACTTTTA-- -CCCATGCAC CCAAATCGCT GCTACAAATT TTTGTTGCAA
Hyphantria_cunea_NPV            AAGAAGTGAT GGAAGTGTTC CTGGTCATTA ACCTGCGT-- -CCCACTCGC CCCAACCGTT GCTACAAATT CTTGGCGCAA
Hyphantria_cunea_NPVN3          AAGAAGTGAT GGACGTGTTC CTGGTCATTA ACCTGCGT-- -CCCACTCGC CCCAATCGTT GCTACAAATT CTTGGCGCAA
HzSNPV_genome                   AAGAAATTAT GGACGTGTTT CTGTCTGTTA ATATGCGA-- -CCAACCAAA CCGAACCGTT GTTACCGATT CTTAGCGCAA
Idaea_virgulata_NPV402          AAGAAGTGAT GGACGTGTTC CTCGTTGTCA ACCTGCGT-- -CCCACTCGC CCCAACCGTT GTTACAAATT CTTGGCGCAA
Junonia_coenia_GV19             AAGTCATCAT GGATCTGTGG TTCGAACTGA GCCTGAAA-- -CCGATGCAG CCGAACCGCT GCTACCAGTT CACGATGCAA
Kotochalia_junodi_NPV205        AAgAGATCAT GGACGTTTTC TTGGTAATCA ATATGAGA-- -CCGACTAGA CCCAACCGTT GCTTTAAATT TTTAGCGCAA
Kotochalia_junodi_NPV32         AAGAGATCAT GGACGTTTTC TTGGTAATCA ATATGAGA-- -CCGACTAGA CCCAACCGTT GCTTTAAATT TTTAGCGCAA
Lacanobia_orecalcea_GVA6        AAGAAATCAT GGACCTGTGG TTCGAGCTGC AGCTTCGA-- -CCAATGCAA CCCAACCGTT GCTACAAGTT TACGATGCAG
Lasiocampa_quercus_NPV202       AGGAGATAAT GGATGTCTTT TTGGTTGTAA ACATGCGC-- -CCTACCAAG CCCAATCGCT GCTACAGATT CTTGGCTCAA
LdMNPV                          AGGAAGTGAT GGACATTTAT CTCACGATCA ACGTGCGC-- -CCCACCAGG CCCAACAGAT GCTACAAGTT CGTGGCGCAG
Lonomia_obliqua_NPV             AAGAAATCAT AGACATTTTC TTAGTGATCA ATCTAAGA-- -CCAACGCGA CCAAACAGAT GTTACAGATT TTTGGCTCAA
Lymantria_monacha_NPVB4         AGGAGGTGAT GGACATCTAT CTGACGATCA ACATGCGA-- -CCCACTAGA CCCAACAGAT GCTACAAATT CGTGGTGCAG
Lymantria_nanayi_NPV750         AAGAAGTGAT GGACATTTAT CTCACTATCA ACGTGCGC-- -CCCACCAGG CCCAACAGAT GCTACAAGTT CGTGGCGCAG
Lymantria_obfuscata_NPV193      AGGAAGTGAT GGACATTTAT CTCACGATCA ACGTGCGC-- -CCCACCAGG CCCAACAGAT GCTACAAGTT CGTGGCGCAG
Mahasena_corbetti_GV699         AAGAAATTAT GGATGTGTTT CTGACAGTCA ACATGCGC-- -CCTACTAAA CCCAACAGAT GTTACCAGTT CTTGGCTCAA
Mahasena_corbetti_NPV595        AAGAAATAAT GGACGTATTC CTCGTGATAA ATATGAGA-- -CCCACTCGT CCTAATAGAT GTTTCCGTTT CCTCGCTCAA
Mahasena_corbetti_NPV692        AAGAAATAAT GGACGTATTC CTCGTGATAA ATATGAGA-- -CCCACTCGT CCTAATAGAT GTTTCCGTTT CCTCGCTCAA
Malacosoma_neustria_NPV         AGGAAATAAT GGACGTGTTT CTAGTGGTCA ACATGCGC-- -CCAACGAAA CCGAACCGTT GTTTCAGATT TTTGGCTCAA
Malacosoma_neustria_NPVk12      AAGAAATAAT GGACGTGTTC CTAGTGGTCA ACATGCGT-- -CCCACGAAA CCTAATCGTT GTTTCAGATT TTTGGCACAA
Malacosoma_pluviale_NPVk3       AAGAAATAAT GGACGTGTTC CTAGCGGTCA ACATGCGT-- -CCACGCAAA CCTAATCGTT GTTTGACATT TTTGGCACAA
Malocosoma_disstria_NPV         AAGAAATAAT GGACGTGTTC CTAGTGGTCA ACATGCGC-- -CCGACGAAG CCGAATCGTT GTTTCAGATT TTTGGCTCAA
Mamestra_brassicae_NPV          AAGAAGTCAT GGACGTTTTC CTTGTAATCA ACATGCGT-- -CCCACTAGA CCCAACCGTT GTTTCAAATT CCTGGCTCAA
Mamestra_brassicae_NPV274       AAGAAGTGAT GGATGTTTTC CTTGTTGTCA ACATGCGT-- -CCCACTAGA CCCAACCGTT GTTACAAATT CCTGGCCCAA
Mamestra_brassicae_NPVK9        AAGAAGTCAT GGACGTTTTC CTTGTAATCA ACATGCGT-- -CCCACTAGA CCCAACCGTT GTTTCAAATT CCTGGCTCAA
Mamestra_configurata_NPVA       AAGAAGTCAT GGATGTTTTC CTCGTAATCA ACATGCGT-- -CCCACTAGA CCCAACCGTT GTTACAAATT CCTGGCGCAG
Mamestra_configurata_NPVB       AAGAAGTCAT GGACGTTTTC CTTGTAATCA ACATGCGT-- -CCCACTAGA CCCAACCGTT GTTTCAAGTT CCTGGCTCAA
Melanchra_persicariaGV26        AAGAAATCAT GGACCTGTGG TTCGAGCTGC AGCTTCGA-- -CCAATGCAA CCCAACCGTT GCTACAAGTT TACGATGCAG
Mynthimna_pseudolatia_NPVA1     AAGAAGTGAT GGATGTCTTC CTTGTCATCA ACATGCGT-- -CCCACTAGA CCCAACCGCT GCTTCAAGTT CCTGGCCCAG
'Mynthimna separata (Lese)NPV'  AAGAAGTCAT GGACGTTTTC CTTGTAATCA ACATGCGT-- -CCCACTAGA CCCAACCGTT GTTTCAAGTT CCTGGCTCAA
Mynthimna_separata_ChinaNPVA3   AGGAGGTGAT GGATGTGTTC CTCGTGGTTA ACATGCGA-- -CCAACCCGT CCCAACCGTT GTTTCAGATT CCTCGCCCAG
Mynthimna_separata_IndiaNPVA4   AAGAGGTGAT GGATGTGTTC CTCGTGGTTA ACATGCGA-- -CCAACCCGT CCCAACCGTT GTTTCAGGTT CCTCGCCCAG
Mynthimna_separata_Japan_NPVA2  AAGAAGTCAT GGACGTTTTC CTTGTAATCA ACATGCGT-- -CCCACTAGA CCCAACCGTT GTTTCAAGTT CCTGGCTCAA
Natada_nararia_GV254            AAGAAATTAT GGACTTGTGG TTTGAAATTC AATTGAGA-- -CCCATGCAA CCTAATCGTT GTTATAAATT TACTCTACAA
Nymphalis_io_NPV353             AAGAAATAAT GGATGTATTC CTTGTGATCA ATATGCGT-- -CCCACTAAA CCTAATAGAT GCTACAAGTT CCTTGCTCAA
Operoptera_brumata_NPV291       AAGAAATTAT GGATGTGTAT TTGACAGTCA ACATGCGC-- -CCTACCAAA CCCAACCGTT GTTACAAGTT TTTGGCTCAA
```

```
Operoptera_brumata_NPVi4        AAGAAATTAT GGATGTGTAT CTGACAGTCA ACATGCGC-- -CCTACCAAA CCCAACCGTT GTTACAAGTT TTTGGCTCAA
Operoptera_brumata_NPVi5        AAGAAATTAT GGATGTGTAT CTGACAGTCA ACATGCGC-- -CCTACCAAA CCCAACCGTT GTTACAAGTT TTTGGCTCAA
OpMNPV                          AAGAAGTAAT GGACGTGTTT CTCGTCGTCA ACATGCGT-- -CCCACGCGC CCCAACCGCT GCTACAAATT CTTGGCGCAA
OpSNPV                          AAGAAATCAT GGACGTGTTT TTAGTTATTA ACATGCGT-- -CCCACCAGG CCCAACCGCT GTTTCAGATT TTTGGCTCAA
Orgya_antiqua_NPVA9             AAGAAATCAT GGACGTGTTT TTAGTTATTA ACATGCGT-- -CCCACCAGG CCCAACCGCT GTTTCAGATT TTTGGCTCAA
Orgya_leucostigma_NPVB6         AAGAAATTAT GGATGTATTC TTGGTGATTA ACATGCGT-- -CCGACAAGA CCTAACCGCT GTTTCAGATT TTTGGCGCAA
Orgya_mixta_NPV67               AGGAAATTAT GGATGTGTTT CTGGTGATCA ACATGCGT-- -CCCACACGA CCAAACAGAT GCTACAGATT TTTGGCGCAA
Orgya_pseudotsugattaNPVB5       AAGAAATAAT GGACGTGTTC TTGGTGATTA ACATGCGT-- -CCCACAAGA CCTAACCGCT GTTTCAAATT CTTGGCGCAA
Orgyia_anartoides_NPV           AAGAAATTAT GGACGTGTTT TTAGTTGATC ACATGCGT-- -CCCACTAGA CCTAACCGCT GCTTTAGATT CCTGGCTCAG
Pachypasa_papyri_NPV436         AAGAAATAAT GGATGTTTTC CTGGTTGTAA ACATGCGT-- -CCCACAAAA CCTAATCGTT GTTACAGATT TTTGGCTCAA
Panolis_flammea_MNPV            AAGAAGTCAT GGATGTTTTC CTCGTAATCA ACATGCGT-- -CCCACTAGA CCCAACCGTT GTTACAAATT CCTGGCGCAG
Panolis_flammea_NPVi3           AAGAAGTCAT GGATGTTTTC CTCGTAATCA ACATGCGT-- -CCCACTAGA CCCAACCGTT GTTACAAATT CCTGGCGCAG
Perina_nuda_NPV                 AAGAGGTGAT GGACGTGTTT CTCGTCGTCA ACATGCGG-- -CCCACGCGC CCAAACCGCT GCTACAAATT TTTGGCCCAA
PhopGV                          AAGAAATTAT GGATCTGTGG TTCGAACTCC AGGTACGT-- -CCCATGCAT CCTAACCGTT GCTACAAATT CACTATGCAA
Phthorimaea_operculella_GVC3    AAGAAATTAT GGATCTGTGG TTCGAACTCC AGGTACGT-- -CCCATGCAT CCTAACCGTT GCTACAAATT CACTATGCAA
Phthorimaea_operculella_GVC4    AAGAAATTAT GGATCTGTGG TTCGAACTCC AGGTACGT-- -CCCATGCAT CCTAACCGTT GCTACAAATT CACTATGCAA
Pieris_brassicae_GV             AACAAATAAT GAACATGTGG TTCGAGATTC AAGTAAGA-- -CCAATGCAA CCTAACCGTT GCTACAAGTT TACTATGCAA
Pieris_brassicae_GV384          AACAAATAAT GGACTTGTGG TTCGAGATTC AATGTAGA-- -CCAATGCAA CCTAACCGTT GCTACAAGTT TACTATGCAA
Pieris_rapaeGV95                AACAAATAAT GGACTTGTGG TTCGAGATTC AATGTAGA-- -CCAATGCAA CCTAACCGTT GCTACAAGTT TACTATGCAA
Plodia_interpunctella_GVB3      AGGAGATCAT GGATCTGTGG TTCGAGCTAC AGCTCCGT-- -CCTATGCAT CCCAATCGCT GTTACAAGTT CACCATGCAG
Plusia_gammaNPV287              AAGAAGTGAT GGATGTTTTC CTTGTTGTCA ACATGCGT-- -CCCACTAGA CCCAACCGTT GTTACAAATT CCTGGCCCAA
Plusia_orichalcea_NPV           AGGAAATTAT GGATGTCTTT CTTGTAATAA ACATGCGC-- -CCAACCAGA CCCAACAGAT GCTTCAAGTT CCTTGCACAG
'Polygonia c-album NPV380'      AAGAAATAAT GGATGTATTC CTTGTGATCA ATATGCGT-- -CCCACTAAA CCTAATAGAT GCTACAAGTT CCTTGCTCAA
PxGV                            AGGAAATTAT GGATCTGTGG TTTGAACTTC AGCTGCGA-- -CCGATGCAG CCCAACCGCT GCTACAAGTT CCTCATGCAG
RaouNPV                         AAGAGGTGAT GGACGTGTTC CTCGTCGTCA ACCTCAGA-- -CCCACACGC CCCAATAGGT GCTACAAGTT CCTCGCTCAA
Samia_ricini_NPV28              AAGAAGTGAT GGATGTTTTC CTTGTTGTCA ACATGCGT-- -CCCACTAGA CCCAACCGTT GTTACAAATT CCTGGCCCAA
Shpinx_ligustri_NPV334          AAGAGGTGAT GGACGTGTTC CTCGTCGTCA ACCTCAGA-- -CCCACACGC CCCAACAGGT GCTACAAGTT CCTCGCTCAA
Smerinthus_ocellata_NPV819      AAGAAGTGAT GGACGTGTTT CTAGTGGTGA ACATGCGT-- -CCCACTAGA CCGAACCGTT GCTTTAGATT TTTGGCGCAA
Sphinx_ligustri_NPV362          AAGAAGTGAT GGATGTTTTC CTTGTTGTCA ACATGCGT-- -CCCACTAGA CCCAACCGTT GTTACAAATT CCTGGCCCAA
Spilosoma_lutea_GV126           AGGAAATCAT GGACTTGTGG TTCGAAATTA AACTGCTG-- -CCAATGCAC CCTAATCGGT GTTACAAATT CTTGTTACAA
Splisoma_imparilis_NPVN2        AAGAAGTGAT GGACGTGTTC CTGGTCATTA ACCTGCGT-- -CCCACTCGC CCCAATCGTT GCTACAAATT CTTGGCGCAA
Spodoptera__exempta_NPVK11      AAGAGATCAT GGACGTTTTC CTAGTTGTAA ATATGCGT-- -CCCACTAAA CCTAACCGTT GTTACAGATT CCTCGCACAA
Spodoptera_exigua_MNPV          AAGAAATCAT GGACGTTTTC CTCGTAATCA ACATGAGA-- -CCAACGAGA CCTAACCGTT GCTTCCGATT TTTGGCTCAG
Spodoptera_frugiperda_NPV       AAGAAATTAT GGACGTGTTC CTAGTGATCA ACATGAGA-- -CCCACTAGG CCCAACAGAT GTTTCAGATT CCTGGCGCAA
Spodoptera_frugiperda_NPVK10    AAGAAATTAT GGACGTGTTC CTAGTGATCA ACATGAGA-- -CCCACTAGG CCCAACAGAT GTTTCAGATT CCTGGCGCAA
Spodoptera_litoralis_NPV        AAGAAGTGAT GGACGTGTTT CTAGTGGTGA ACATGCGT-- -CCCACTAGA CCGAACCGTT GCTTTAGATT TTTGGCGCAA
Spodoptera_littoralis_GV66      AAGAAATTAT GGACTTGTGG TTTGAGCTCA CACTGAGA-- -CCCATGCAG CCTAACCGTT GCTACCGTTT CACCATGCAA
Spodoptera_littoralis_NPV440    AAGAAGTGAT GGACGTGTTT CTAGTGGTGA ACATGCGT-- -CCCACTAGA CCGAACCGGT GCTTTAGATT TTTGGAGCAA
Spodoptera_litura_GV766         AAGAAATTAT GGACTTGTGG TTTGAGCTCA CACTGAGA-- -CCCATGCAG CCTAACCGTT GCTACCGTTT CACCATGCAA
Spodoptera_litura_NPVg          AAGAAATAAT GGACGTCTTC CTAGTCATCA ACATGAGA-- -CCCACTAGA CCCAACCGAT GCTACAGATT CTTGGCCCAA
Spodoptera_litura_NPV2          AAGAAGTGAT GGACGTGTTT TTAGTGGTAA ACATGCGT-- -CCCACTAGA CCGAACCGTT GCTTTAGATT TTTAGCGCAG
Spodoptera_sunia_NPVK8          AGGAAATCAT GGACGTGTTC CTTGTGGTCA ACATGCGC-- -CCGACGAGA CCTAACAGAT GCTTTAAATT CCTAGCTCAA
Spodoptera_sunia_NPVK7          AGGAAATTAT GGACGTTTTT CTCctTTTTC AGATGAGA-- -CCCACGAGA CCTAACCGAT GCTTCAGATT TTTAGCTCAG
Thysanoplusia_orichalcea_NPV    AAGAGGTTAT GGACGTGTTC CTTGTCGTGA ACCTTAGA-- -CCCACGCGC CCCAACAGGT GCTACAAATT CCTTGCCCAG
Thysanoplusia orichalcea NPV2   AgGAaaTTAT GGAtGTcTTt CTTGTaaTaA ACatgcGc-- -CCaACcaGa CCCAACAGaT GCTtCAAgTT CCTTGCaCAG
Trabala_vishnou_NPV195          AAGAAATTAT GGACGTGTTT TTGGTGATAA ACATGCGA-- -CCCACTAGA CCCAACAGAT GCTTTAGGTT CTTGGCCCAG
Trichoplusia_ni_GV              AGGAAATTAT GGATCTGTGG TTTGAGCTGC AACTTCGG-- -CCCATGCAG CCCAACAGGT GCTACAAGTT CACGATGCAA
Trichoplusia_ni_MNPVC2          AAGAAGTGAT GGATGTTTTC CTTGTTGTCA ACATGCGT-- -CCCACTAGA CCCAACCGTT GTTACAAATT CCTGGCCCAA
Trichoplusia_ni_SNPV            AAGAAATCAT GGACGTATTT CTTGTGGTTA ACATGCGC-- -CCGACAAGA CCCAATCGTT GCTTCAAATT CTTAGCCCAA
Trichoplusia_orichalceaNPVB9    AAGAAATCAT GGACGTGTTC CTCGTAGTGA ACATGCGT-- -CCCACTAGG CCCAACCGTT GCTTCAAATT CCTAGCCCAA
Utetheisa_pulchella_NPV142      AAGAAATTAT GGATGTGTAT CTGACATTCA ACATGCGC-- -CCTACCAAA CCCAACCGGT GCTACAAGTT TTTGGCTCAA
Vanessa_alatanta_NPV54          AAGAAATAAT GGATGTATTC CTTGTGATCA ATATGCGT-- -CCCACTAAA CCTAATAGAT GCTACAAGTT CCTTGCTCAA
Vanessa_cardui_NPV367           AAGAAATAAT GGATGTATTC CTTGTGATCA ATATGCGT-- -CCCACTAAA CCTAATAGAT GCTACAAGTT CCTTGCTCAA
Wiseana_cervinata_GV342         AAGAAATAAT GGATCTGTGG CTTAGAATTG CAACATTCGT CCCCATGCAG CCCAACCGCT GTTACAGGTT TACTATGCAA
Wiseana_cervinata_NPV344        AAGAGGTGAT GGACGTCTTC CTCGTAATTA ACATGCGA-- -TCGACTAAA CCCAATCGTT GTTTCAGATT TCTAGCTCAG
Wiseana_signata_NPV             AAGAGGTGAT GGACGTCTTC CTCGTAATTA ACATGCGA-- -TCGACTAAA CCCAATCGTT GTTTCAGATT TCTAGCTCAG
XcGV                            AAGAAATTAT GGACCTGTGG TTTGAACTGC AACTGAGA-- -CCCATGCAG CCCAACAGAT GTTACAAGTT TACCATGCAA

Abraxasgrossulariata_NPV112     CACGCG---C TCAGGTGGGA CTGCGACTAC GTGCCCCACG AGGTGATCAG AATCGTTGAG CCATCGTACG TGGGCATGAA
Achaea_faber_NPV64              CACGCG---C TCCGTTGCGA TCCCGACTAC GTGCCCCATG AAGTCATCCG CATCGTCGAG CCCGTGTACG TAGGAAACAA
AchaeajanataGV835               TACGCG---C TGcGCCGCCA ACCCGAATAC GTTGCACACG ACGTGATTCG TCAACAGGAC CCCTACTACA TTGGTCCCGA
AcMNPV                          CACGCT---C TGCGTTGCGA CCCCGACTAT GTACCTCATG ACGTGATTAG GATCGTCGAG CCTTCATGGG TGGGCAGCAA
AcMNPVK1                        CACGCT---C TGCGTTGCGA CCCCGACTAT GTACCTCATG ACGTGATTAG GATCGTCGAG CCTTCATGGG TGGGCAGCAA
Actias_selene_NPV47             CACGCG---C TCAGATGGGA CTGCGACTAC GTGCCGCACG AGGTAATCCG CATTGTGGAG CCATCCTACG TGGGCATGAA
Adoxophyes_orana_GV             TACGCG---C TGGGCGCTCA CCCCGATTAC GTGGCTCACG ATGTTATCCG TCAACACGAT CCTTACTACG TAGGACCCAA
AfMNPVK2                        CACGCT---C TTATATGGGA CGAAGACTAC GTGCCCCACT tAGTAATCAG AATTGCGGAG CCATCTTACG TGGGTATGAA
Aglais_urticae_NPVA7            CACGCT---C TTCGTTGCGA TCCCGATTAC GTTCCTCACG AAGTCATCCG CATCGTCGAA CCATCATACG TAGGCAGCAA
Agraulis_vanillare_NPV779       CACGCG---C TGCGGTGGGA YTGCGATTAY GTGCCCCACG ARGTGATCAG RATCGTGGAG CCCTCRTACG TGGGCATGAA
Agrotis_segetum_NPVi6           CACGCT---C TGCGCTGCGA TCCCGATTAC GTGCCCCACG AGGTGATCCG CATCGTCGAA CCTGTGTACG TCGGTGGCAA
Agrotis_segetum_NPVK4           CACGCT---A TGCGCTGCGA TCCCGATTAC GTGCCCCACG AGGTGATCCG CATAGTCGAA CCTGTGTACG TTGGTGGCAA
Ammalo_insulata_NPV175          CACGCG---C TCAGGTGGGA CTGTGACTAC GTGCCGCACG AAGTCATTCG CATTGTAGAG CCCTCGTACG TGGGCATGAA
Amonis_sabulifera_NPV14         CACGCT---C TTCGTTGCGA AGACGACTAT GTGCCTCACG AGGTCATCAG AATTGTGGAA CCCTCGTACG TAGGAGATCG
Amsacta_albistriga_NPV          CACGCG---C TTCGTTGGGA CGACAATTAT GTTCCCCACG AAGTGATTAG GATTGTGGAG CCATCGTACG TGGGCATGAA
Antherea_pernyi_NPV             CACGCG---C TCAGATGGGA CTGCGATTAC GTGCCGCACG AGGTAATCCG CATTGTGGAG CCATCCTACG TGGGCATGAA
Antherea_polyphemus_NPV30       CACGCT---C TTAGGTGGGA CGAAGACTAC GTGCCCCACG AAGTAATCAG AATTGTGGAG CCATCCTACG TGGGCATGAA
Anthophila_fabricana_GV790      TACGCT---C TCTGCGCCAA TCCCGACTAT GTCGCACATG ATGTCATTCG TCAACAAGAT CCCTACTATG TCGGTTCCAA
Anticarsia_gemnitalis_NPV       CACGCT---C TCCGTTGGGA CTGCGATTAC GTGCCCCACG AGGTAATCCG CATTGTGGAG CCTTCCTACG TGGGCATGAA
Aporia_crataegi_NPV106          CACGCT---C TTCGTTGGGA TTGCGACTAT GTGCCCCACG AAGTTATCAG AATTGTTGAA CCGTCTTACG TCGGTATGAA
Archips_cerasivora_NPV          CACGCG---C TCAGGTGGGA CTGCGATTAT GTGCCCCACG AAGTGATTAG AATCGTTGAG CCATCGTACG TAGGTATGAA
Aroa_discalis_NPV63             CACGCT---C TGCGTTGCGA CCCAGAATAC GTGCCCCACG AGGTTATCCG TATCGTAGAA CCCAGCTACG TGGGCAGCAA
Artogeia_napi_GV242             TACGCT---T TAGGTGCTCA CCCCGACTAC GTTCCCCATG ACGTTATTCG CGCCCAGGAT CCTTACTACA TTGGTCCCAA
Attacus_ricini_NPV              CACGCC---G TTAGATGGGA CTGCGACTAC GTGCCGCACG AGGTAATCCG CATTGTGGAG CCATCCCACG TGGGCATGAA
BmNPV                           CACGCT---C TTAGGTGGGA AGAAGACTAC GTGCCCCACG AAGTAATCAG AATTGTGGAG CCATCCTACG TGGGCATGAA
Bombyx_mori_NPV460              CACGCT---C TTAGGTGGGA CGAAGACTAC GTGCCCCACG AAGTAATCAG AATTATGGAG CCATCCTACG TGGGCATGAA
Buzura_edwardsii_NPV505         CACGCT---C TGCGTTGCGA CCCTGACTAC GTTCCTCACG AAGTCATTCG CATTGTGGAG CCCGTGTGCG TAGGAAACAA
Buzura_Supressaria_NPV          CACGCG---C TCCGTTGCGA CCCCGACTAC GTGCCACACG AGGTGATCCG AATTGTCGAG CCCAGCTACG TAGGTAGCAA
Cacoecia_murinana_GV37          TACGCT---C TTGGTGCTAA CCCCGATTAC GTTGCACATG ATGTAATCCG TCAACAGGAT CCTTACTACG TTGGACCTAA
Catopsila_pomona_NPV78          CACGCA---C TGCGTTGGGA TTGCGACTAT GTGCCGCACG AAGTTATCAG AATTGTTGAA CCATCATACG TGGGCATGAA
Cerapteryx_gramini_NPVA5        CACGCT---C TCMGATGCGA YCCAGACTAC GTGCCCCACG ATGTGATCCG CATCGTYGAG CCYTTGTACG TCGGCTCCAA
ChfuNPV                         ---CCC---C TCAGGTGGGA CTGCGATTAC GTGCCGCACG AGGTGATTAG AATCGTTGAG CCATCGTACG TGGGTATGAA
ChfuNPVDEF                      CACGCT---C TCCGCTGGGA CTGCGATTAC GTGCCCCACG AAGTAATCCG CATTGTGGAG CCTTCCTACG TGGGCATGAA
Choristoneura_furamifera_GV     TACGCG---C TTGGTGCTAA CCCCGAATAC GTTGCACATG ATGTAATCCG TCAACAGGAT CCTTACTACG TTGGTCCTAA
Choristoneura_rosaceana_NPV     CACGCG---C TCAGGTGGGA CTGCGATTAC GTGCCGCACG AGGTGATTAG AATCGTTGAG CCATCGTACG TGGGTATGAA
Cirphis_unipuncta_NPV172        CATGCT---C TGCGTTGCGA TCCCGACTAC GTGCCCCACG AAGTCATTCG CATCGTTGAG CCATCGTACG TGGGCAGCAA
CpGV                            TACGCT---C TCGGCGCCCA CCCCGATTAT GTTGCTCACG ACGTGATCCG CCAACAAGAC CCCTACTACG TGGGCCCAAA
Cryptophlebia_leucotreta_GV     TATGCT---C TCTGTGCCCA TCCCGATTAT GTCGCTCACG ATGTGATCCG CCAGCAGGAT CCCTACTATG TAGGACCTAA
Cynosarga_chrysolopa_NPV302     CACGCG---C TCAGATGGGA CTGCGATTAC GTGCCCCACG AAGTCATCAG AATCGTGGAG CCGTCTTACG TGGGCATGAA
Darna_trima_GV545               TACGCT---C TAGGTGCGCA TCCCGATTAC GTAGCTCATG ATGTTATTCG TCAACAAGAT CCTTATTACG TTGGACCTAA
Diacrisia_obliqua_NPV798        CATGCC---C TCCGGTGGGA CGACAATTAC GTGCCCCACG AGGTCATTAG GATTGTTGAA CCTTCGTACG TAGGCATGAA
Diatrea_saccharalis_NPVK6       TACGCA---C TAGGAGCCAA TCCAGATTAT ATCGCACATG ACGTTATCAG ACAACAAGAC CCCTATTATG TGGGACCTAA
Ectropis_obliqua_NPV            CACGCG---C TGCGTTGCGA TCCCGATTAC GTGCCGCACG AGGTGATCAG AATCGTTGAG CCCAGCTACG TGGGCAGCAA
Ephestia_cautella_NPVC5         CACGCT---C TCCGATGGGA TGACGATTAT GTGCCTCATG AGGTGATTCG AGTAATCGAG CCGTCATATG TAGGCATGAA
EppoMNPV                        CACGCT---C TCAGATGGGA TTGCGATTAC GTGCCCCACG AGGTGATAAG AATTGTGGAA CCTTCTTACG TGGGCATGAA
Erannis_defolariaNPV174         CACGCT---C TGCGTTGCGA CGCCGATTAC GTGCCTCATG AGGTGATTCG CATTGTCGAG CCCAGTTATG TGGGTAGCAA
```

```
Euplasia_luciparaGV248              TACGCC---C TGTGCGCGCA CCCCGACTAC GTCGCTCACG ATGTGATCCG CCAACAGGAC CCTTACTACG TGGGCCCCAA
Euproctis_Chrysorrohea_NPVA10       CACGCT---T TGCGTTGGGA GTGCGATTAC GTGCCCCACG AAGTCATTCG TATTGTGGAG CCGTCATACG TAGGCATGAA
Euproctis_similis_NPV768            CACGCT---T TGCGTTGGGA GTGCGATTAC GTGCCCCACG AAGTCATTCG TATTGTGGAG CCGTCATACG TAGGCATGAA
Harrisina_brillians_GV              TATGCG---C TCTGTACTCA TCCTGATTAT ATCGCTCATG ATGTAATCCG TCAACAAGAT CCCTATTTTG TGGGACCCAA
Harrisina_brillians_GVM2            TATGCG---C TCTGTACTCA TCCTGATTAT ATCGCTCATG ATGTAATCCG TCAACAAGAT CCCTATTTTG TGGGACCCAA
HaSNPV                              CACGCT---C TGCGTTGTGA TCCCGACTAT ATTCCTCACG AAGTCATTCG TATTGTAGAA CCTTCCTATG TAGGCAGTAA
Heliconius_erato_NPV789             CACGCG---C TCCGGTGGGA CTGCGATTAC GTGCCCCACG AGGTGATCAG GATCGTGGAG CCCTCGTACG TGGGCATGAA
Homona_coffearia_GV744              TACGCG---C TTGGCGCCAA CCCAGATTAC GTTGCACATG ATGTAATTCG TCAACAGGAT CCTTACTACG TTGGTCCTAA
Homona_coffearia_NPV745             TACGCG---C TTGGCGCCAA CCCAGATTAC GTTGCACATG ATGTAATTCG TCAACAGGAT CCTTACTACG TTGGTCCTAA
Hyphantria_cunea_GV228              TATGCT---C TATGTGCACC GCCAGGCTAC ATCGCTCACG ATGTTATAAG AGCACAAGAT CCTTATTATG TAGGATCCGA
Hyphantria_cunea_NPV                CATGCT---C TCCGGTGGGA CGACAATTAC GTGCCCCACG AGGTCATTAG GATTGTTGAA CCTTCGTACG TAGGCATGAA
Hyphantria_cunea_NPVN3              CATGCC---C TCCGGTGGGA CGACAATTAC GTGCCCCACG AGGTCATTAG GATTGTTGAA CCTTCCTACG TAGGCATGAA
HzSNPV_genome                       CACGCT---C TGCGTTGTGA TCCCGACTAT ATTCCTCACG AAGTCATTCG TATTGTAGAA CCTTCCTATG TAGGCAGTAA
Idaea_virgulata_NPV402              CACGCA---C TTAGGTGGGA CTGCGATTAC GTGCCCCATG AGGTGATCCG CATCGTGGAG CCATCTTACG TGGGCATGAA
Junonia_coenia_GV19                 TACGCT---C TAGCCGCCAA CCCTGATTAC GTAGCGCCCG ACGTGATTCG TCAGCAAGAC CCCTATTACG TCGGTCCTGA
Kotochalia_junodi_NPV205            CACGCT---C TTCGTTGTGA TCCCGATTAT GTGCCTCACG AGGTTATACG TATAGTCGAA CCCTCGTATG TGGGTAGCAA
Kotochalia_junodi_NPV32             CACGCT---C TTCGTTGTGA TCCCGATTAT GTGCCTCACG AGGTTATACG TATAGTCGAA CCCTCGTATG TGGGTAGCAA
Lacanobia_orecalcea_GVA6            TACGCT---C TGGGCGCCAA CCCCGACTAC GTCGCTCATG ACGTAATCCG CCAACAAGAC CCCTACTACG TGGGTCCCAA
Lasiocampa_quercus_NPV202           CACGCG---C TGCGTTGCGA CCCTGATTAT GTTCCGCATG AGACGATCAG AATCGTGGAG CCGTCGTACG TGGGCAGCAA
LdMNPV                              CACGCC---C TGCGGTGCGA CGAGGGCTAC GTGCCGCACG AGGTGATCCG CATCGTGGAG CCCTCCACGG TCGAG---AA
Lonomia_obliqua_NPV                 CACGCT---C TTCGTTGGGA CTGCAATTAT GTACCTCACG AAGTGATTAG AATCGTCGAG CCTTCTTACG TGGGCACGAA
Lymantria_monacha_NPVB4             CACGCT---C TGCGTTGCGA CCCCGACTAC GTGCCGCACG AAGTGATCCG CATCGTAGAG CCCGCCACCG TT---GAGAA
Lymantria_nanayi_NPV750             CACGCC---C TGCGGTGCGA CGAAGGCTAC GTGCCGCACG AAGTAATCCG CATTGTGGAG CCCTCCACGG TCGAG---AA
Lymantria_obfuscata_NPV193          CACGCC---C TGCGGGGCGA CGAGGGCTAC GTGCCGCATG AGGTGATCCG CATCGTGGAG CCCTCCACGG TCGAG---AA
Mahasena_corbetti_GV699             CACGCT---T TGAGATGCGA TCCTGATTAT GTTCCTCATG AGGTGATCAG AATCGTTGAG CCCtGCTGGG TTGGAAGCAA
Mahasena_corbetti_NPV595            CACGCC---T TGAGATGCGA TCCTGATTAT GTTCCTCATG AAGTGATCAG AATCGTTGAA CCGTCATATG TAGGAAGTAA
Mahasena_corbetti_NPV692            CACGCC---T TGAGATGCGA TCCTGATTAT GTTCCTCATG AAGTGCTCAG AATCGTTGAA CCGTCATATG TAGGAAGTAA
Malacosoma_neustria_NPV             CACGCG---T TACGGTGCGA CTCTGACTAT GTCCCGCACG AGGTGATCAG GATTGTGGAA CCTTCGTACG TAGGTAGCAA
Malacosoma_neustria_NPVk12          CACGCG---C TACGGTGCGA CTCTGACTAT GTCCCGCACG AGGTGATCAG AATCGTGGAA CCATCGTACG TAGGCAGCAA
Malacosoma_pluviale_NPVk3           CACGCG---C TACTGAGCGA CTCTGACTAT GTCCCGCACG AGGTGATCAG AATCGTGGAA CCATCGTACG TAGGCAGCAA
Malocosoma_disstria_NPV             CACGCG---C TGCGGTGCGA CTCTGATTAT GTTCCACACG AAGTGATCAG AATCGTGGAG CCTTCGTACG TAGGCAGCAA
Mamestra_brassicae_NPV              CATGCT---C TGCGTTGCGA TCCCGACTAT GTGCCCCACG AAGTCATCCG CATCGTTGAA CCGTCCTACG TGGGCAGCAA
Mamestra_brassicae_NPV274           CACGCT---C TGCGTTGCGA TCCCGACTAT GTGCCTCATG ACGTGATTAG CATCGTTGAG CCTTCATGGG TGGGCAGCAA
Mamestra_brassicae_NPVK9            CATGCT---C TGCGTTGCGA TCCCGACTAT GTGcCCCACG AaGTCATCCG CATCGTTGAA CCGTCCTACG TGGGCAGCAA
Mamestra_configurata_NPVA           CATGCT---C TGCGTTGCGA TCCCGACTAC GTGCCCCACG AAGTCATTCG CATCGTTGAG CCATCATACG TGGGCAGCAA
Mamestra_configurata_NPVB           CATGCT---C TGCGTTGCGA TCCCGACTAT GTGCCCCACG AAGTCATCCG CATCGTTGAA CCGTCCTACG TGGGCAGCAA
Melanchra_persicariaGV26            TACGCT---C TGGGTGCCAA CCCCGACTAC GTCGCTCACG ACGTGATCCG TCAACAAGAC CCCTACTACG TGGGACCCAA
Mynthimna_pseudolatia_NPVA1         CACGCT---C TGCGTTGCGA CCCCGACTAT GTACCGCACG AAGTCATTCG CATTGTCGAG CCTTCGTACG TGGGAAGCAA
'Mynthimna separata (Lese)NPV'      CATGCT---C TGCGTTGTGA TCCCGACTAC GTGCCCCACG AAGTCATCCG CATCGTTGAA CGGTCCTGCG TGGGCACGAA
Mynthimna_separata_ChinaNPVA3       CACGCG---C TCCGCTGCGA TCCCGACTAC GTGCCCCACG ATGTGATCCG CATCGTTGAG CCTTCGTACG TCGGCACCAA
Mynthimna_separata_IndiaNPVA4       CACGCG---C TCCGCTGCGA TCCCGACTAC GTGCCCCACG ATGTGATCCG CATCGTTGAG CCTTCGTACG TCGGCACCAA
Mynthimna_separata_Japan_NPVA2      CATGCT---C TGCGTTGCGA TCCCGACTAC GTGCCCCACG AAGTCATCCG CATCGTTGAA CCGTCCTACG TGGGCAGCAA
Natada_nararia_GV254                TACGCA---T TAGGCGCTCA CCCMGATTAC GTCGCTCATG ATGTTATCCG CCAACAAGAT CCTTATTATG TTGGACCAAA
Nymphalis_io_NPV353                 CACGCT---C TTCGTTGCGA TCCCGATTAC GTTCCTCACG AAGTCATCCG CATCGTCGAA CCATCATACG TAGGCAGCAA
Operoptera_brumata_NPV291           CACGCT---C TGCGTTCAGA TCCCGACTAT GTGCCGCATG AGGTGATCCG TATCGTCGAG CCCAGCTGGG TTGGTAGCGA
Operoptera_brumata_NPVi4            CACGCT---C TGCGTTCAGA TCCCGATTAT GTGCCGCATG AGGTGATCCG TATAGTCGAG CCCACGTGGG TTGGTAGCCA
Operoptera_brumata_NPVi5            CACGCT---C TGCGTTCAGA TCCCGATTAT GTGCCGCATG AGGTGATCCG TATCGTCGAG CCCAGCTGGG TTGGTAGCGA
OpMNPV                              CACGCT---C TCAGGTGGGA CTGCGACTAC GTGCCCCACG AGGTGATTAG GATTGTCGAG CCGTCGTACG TGGGCATGAA
OpSNPV                              CACGCT---T TGCGTTGCGA TCCCGAGTAC GTTCCCCACG AAGTAATTCG CATTGTAGAG CCCAGCTACG TAGGCAGCAA
Orgya_antiqua_NPVA9                 CACGCT---T TGCGTTGCGA TCCCGAGTAC GTTCCCCACG AAGTAATTCG CATTGTAGAG CCCAGCTACG TAGGCAGCAA
Orgya_leucostigma_NPVB6             CACGCT---C TGCGTTGCGA TCCTGAATAC GTGCCTCATG AAGTCATTCG CATAGTAGAA CCTAGTTACG TGGGAAACAA
Orgya_mixta_NPV67                   CACGCT---C TGCGTTGCGA CCCAGAATAC GTGCCCCACG AGGTTATCCG TATCGTAGAA CCCAGCTACG TGGGCAGCAA
Orgya_pseudotsugattaNPVB5           CACGCT---C TGCGTGGGAC TCCTGACTAC GTGCCTCATG AAGTCATTAG GATAGTAGAA CCTACTTACG TGGGAATCAA
Orgyia_anartoides_NPV               CACGCT---T TGCGTTGCGA CGCCGATTAC GTACCTCACG AAGTAATTCG CATCGTGGAG CCCAGTTACG TTGGCAGCAA
Pachypasa_papyri_NPV436             CACGCT---C TTCGTTGTGA TTCAGATTAC GTGCCCCATG AAACGATCAG AATAGTAGAA CCTTCGTATG TAGGCAGCAA
Panolis_flammea_MNPV                CATGCT---C TGCGTTGCGA TCCCGAATAC GTGCCCCACG AAGTCATTCG CATCGTTGAG CCATCATACG TGGGCAGCAA
Panolis_flammea_NPVi3               CATGCT---C TGCGTTGCGA TCCCGAATAC GTGCCCCATG AAGTCATTCG CATCGTTGAG CCATCATACG TGGGCAGCAA
Perina_nuda_NPV                     CACGCT---C TTAGGTGGGA CTGCGATTAC GTGCCCCACG AGGTTATTAG GATTGTCGAG CCGTCGTACG TGGGCATGAA
PhopGV                              TACGCT---C TCGGCGCCCA CCCCGATTAC GTGGCTCATG ATGTGATTCG TCAGCAAGAT CCCTACTACA TTGGTCCCAA
Phthorimaea_operculella_GVC3        TACGCT---C TCGGCGCCCA CCCCGATTAC GTGGCTCATG ATGTGATTCG TCAGCAAGAT CCCTACTACA TTGGTCCCAA
Phthorimaea_operculella_GVC4        TACGCT---C TCGGCGCCCA CCCCGATTAC GTGGCTCATG ATGTGATTCG TCAGCAAGAT CCCTACTACA TTGGTCCCAA
Pieris_brassicae_GV                 TACGCT---T TAGATGCTCA CCCCGACTAC GTTCCCCATG ACGTTATTCG CGCCCAGGAT CCTTACTACA TTGGTCCCAA
Pieris_brassicae_GV384              TACGCT---T TAGGTGCTCA CCCCGACTAC GTTCCCCATG ACGTTATTCG CGCCCAGGAT CCTTACTACA TTGGTCCCAA
Pieris_rapaeGV95                    TACGCT---T TAGGTGCTCA CCCCGACTAC GTTCCCCATG ACGTTATTCG CGCCCAGGAT CCTTACTACA TTGGTCCCAA
Plodia_interpunctella_GVB3          TACGCG---C TCTCTGCCCA TCCCGAGTAC GTCGCCCACG ACGTGATTCG CCAACAGGAT CCGTACTATG TGGGTCCAAA
Plusia_gammaNPV287                  CACGCT---C TGCGTTGCGA CCCCGACTAT GTACCTCATG ACGTGATTAG GATCGTCGAG CCTTCATGGG TGGGCAGCAA
Plusia_orichalcea_NPV               CATGCT---C TGCGTTGCGA CCCTGACTAC GTTCCTCATG AAGTCATCCG CATTGTCGAG CCTTCTTGGG TGGGCAGCAA
'Polygonia c-album NPV380'          CACGCT---C TTCGTTGCGA TCCCGATTAC GTTCCTCACG AAGTCATCCG CATCGTCGAA CCATCATACG TAGGCAGCAA
PxGV                                TACGCT---C TAGGCGCCGA CGCCGATTAC GTTCCTCACG ACGTCATTCG CGCTCAGGAC CCTTACTACG TCGGACCGCA
RaouNPV                             CACGCT---C TTAGATGGGA CGAAGACTAC GTGCCCCACG AAGTAATCAG AATTGTGGAG CCATCTTACG TGGGTATGAA
Samia_ricini_NPV28                  CACGCT---C TGCGTTGCGA CCCCGACTAT GTACCTCATG ACGTGATTAG GATCGTCGAG CCTTCATGGG TGGGCAGCAA
Shpinx_ligustri_NPV334              CACGCT---C TTAGGTGGGA CGAAGACTAC GTGCCCCACG AAGTAATCAG AATTGTGGAG CCATCCTACG TGGGCATGAA
Smerinthus_ocellata_NPV819          CACGCG---C TCCGATGCGA CCCCGAGTAC GTTCCCCACG ACGTGATCCG CATCGTCGAA CCGTCGTACG TCGGCACCAA
Sphinx_ligustri_NPV362              CACGCT---C TGCGTTGCGA CCCCGACTAT GTACCTCATG ACGTGATTAG GATCGTCGAG CCTTCATGGG TGGGCAGCAA
Spilosoma_lutea_GV126               TACGCA---C TATGCGCACC ACCAGGCTAC ATTGCTCATG ACGTCATTCG AGCTCAAGAT CCCTATTATG TGGGACCCGA
Splisoma_imparilis_NPVN2            CATGCC---C TCCGGTGGGA CGACAATTAC GTGCCCCACG AGGTCATTAG GATTGTTGAA CCTTCCTACG TAGGCATGAA
Spodoptera__exempta_NPVK11          CACGCG---C TTCGTTGCGA TCCCGACTAC GTGCCCCATG AAGTCATCCG CATCGTCGAG CCCGTGTACG TAGGAAACAA
Spodoptera_exigua_MNPV              CACGCT---C TCCGTTGCGA TCCCGACTAC GTTCCCCACG AAGTCATCCG CATCGTCGAG CCCGTGTACG TCGGCACCAA
Spodoptera_frugiperda_NPV           CACGCT---C TCCGTTGCGA CCCTGACTAC GTTCCTCACG AAGTCATTCG CATTGTGGAG CCCGTGTACG TAGGAAACAA
Spodoptera_frugiperda_NPVK10        CACGCT---C TCCGTTGCGA CCCTGACTAC GTTCCTCACG AAGTCATTCG CATTGTGGAG CCCGTGTACG TAGGAAACAA
Spodoptera_litoralis_NPV            CACGCG---C TCCGATGCGA CCCCGAGTAC GTTCCCCACG ACGTGATCCG CATCGTCGAA CCGTCGTACG TCGGCACCAA
Spodoptera_littoralis_GV66          TACGCG---T TGGCGGCCAA TCCTGATTAC TGTGCTCACG ACGTTATCCG CCAACAGGAT CCCTATTACG TAGGTCCAGA
Spodoptera_littoralis_NPV440        CACGCG---C TCCGATGCGA CCCCGAGTAC GTTCCCCACG ACGTGATCCG CATCGTCGAA CCGTCGTACG TTGGCACCAA
Spodoptera_litura_GV766             TACGCG---T TGGCGGCCAA TCCTGATTAC TGTGCTCACG ACGTTATCCG CCAACAGGAT CCCTATTACG TAGGTCCAGA
Spodoptera_litura_NPVg              CACGCT---C TCCGTTGCGA TCCTGACTAC GTTCCTCACG AAGTGATCCG CATCGTCGAG CCGTCATACG TCGGCTCCAA
Spodoptera_litura_NPV2              CACGCG---C TCCGATGCGA TCCCGAGTAC GTTCCTCACG ACGTGATCCG TATCGTAGAA CCGTCGTACG TCGGCACCAA
Spodoptera_sunia_NPVK8              CACGCT---C TGCGTTGCGA CCCCGACTAC GTTCCTCATG AGGTGATCAG AATCGTAGAG CCTTCCTGGG TGGGCAGCAA
Spodoptera_sunia_NPVK7              CACGCT---C TTCGGTGGGA TCCCGATTAC GTTCCTCACG AGGTCATCCG TATCATTGAG CCGGCGTACG TCGGTGCCAA
Thysanoplusia_orichalcea_NPV        CACGCC---C TCAGGTGGGA CTGCGATTAC GTGCCCCACG AAGTTATCAG AATTGTGGAA CCTTCCTACG TTGGCATGAA
Thysanoplusia orichalcea NPV2       CAtGCt---C TgcGttgcGA CcctGAcTAC GTtCCtCAtG AAGTcATCcG cATTGTcGAg CCTTCtTggG TgGGCAgcAA
Trabala_vishnou_NPV195              CACGCG---C TTAGATGCGA TCCTGACTAT GTGCCTCACG AAGTGATTCG CATTGTCGAA CCCAGCTACG TGGGCAGCAA
Trichoplusia_ni_GV                  TACGCC---C TGGCTGCGAA CCCCGACTAC GTCGCTCACG ATGTGATTCG CCAACACGAT CCCTACTACG TCGGACCCGA
Trichoplusia_ni_MNPVC2              CACGCT---C TGCGTTGCGA CCCCGACTAT GTACCTCATG ACGTGATTAG GATCGTCGAG CCTTCATGGG TGGGCAGCAA
Trichoplusia_ni_SNPV                CACGCT---T TACGTTGCGA CCCCGATTAT GTTCCTCACG AGGTGATTAG AATCGTAGAG CCGTCTTGGG TAGGCAACAA
Trichoplusia_orichalceaNPVB9        CACGCT---C TGCGTTGCGA CCCCGACTAT GTTCCTCATG AAGTGATCAG AATCGTTGAG CCTTCATGGG TCGGCAGCAA
Utetheisa_pulchella_NPV142          CACGCT---C TGCGTTCAGA TCCCGACTAT GTGCCGCATG AAGTGATCCG TATCGTCGAG CCCTCCTAcG TTGGTAGCCA
Vanessa_alatanta_NPV54              CACGCT---C TTCGTTGCGA TCCCGATTAC GTTCCTCACG AAGTCATCCG CATCGTCGAA CCATCATACG TAGGCAGCAA
Vanessa_cardui_NPV367               CACGCT---C TTCGTTGCGA TCCCGATTAC GTTCCTCACG AAGTCATCCG CATCGTCGAA CCATCATACG TAGGCAGCAA
Wiseana_cervinata_GV342             TACGCCGCTT TGTACTCATC CCGAATATAT TGCCGCCACG ACGTAATCCG TCAGCAAGAT CCTATTATCG TAGGCTCCAA
Wiseana_cervinata_NPV344            CACGCG---C TTCGCTGCGA CCCCGACTAT GTGCCGCACG AAATCATCCG CATAGTGGAG CCGTCGTACG TGGGC---AA
Wiseana_signata_NPV                 CACGCG---C TTCGCTGCGA CCCCGACTAT GTGCCGCACG AAATCATCCG CATAGTGGAG CCGTCGTACG TGGGC---AA
XcGV                                TACGCT---C TGGCTGCCAA CCCCGATTAC GTTGCCCACG ACGTGATCCG TCAGCACGAT CCCTATTACG TTGGTCCCGA
```

```
Abraxasgrossulariata_NPV112     CAACGAGTAC CGCATCAGCC TGGCAAAGAA GGGCGGCGGC TGCCCCATCA TGAACATCCA CTCCGAATAC ACCAATTCAT
Achaea_faber_NPV64              CAACGAGTAC CGCATCAGCC TGGCCAAGAA GGGCGGCGGT TGCCCCGTGA TGAACCTGCA TTCCGAGTAC ACCAACAGCT
AchaeajanataGV835               CAACAAGGAG CGCATCAACC TGTCCAAATC TGCCTACGCG TTCCCGTTGA CCTGTCTCCA GTCGGTGTAC AACGAGAACT
AcMNPV                          CAACGAGTAC CGCATCAGCC TGGCTAAGAA GGGCGGCGGC TGCCCAATAA TGAACCTTCA CTCTGAGTAC ACCAACTCGT
AcMNPVK1                        CAACGAGTAC CGCATCAGCC TGGCTAAGAA GGGCGGGGGC TGCCCAATAA TGAACCTTCA CTCTGAGTAC ACCAACTCGT
Actias_selene_NPV47             CAACGAGTAC AGAATTAGCC TCGCCAAGAA AGGCGGCGGC TGCCCCATCA TGAACATTCA CAGCGAGTAC ACCAACTCGT
Adoxophyes_orana_GV             CAACATTGAA CGTATCAATC TGAGCAAGAA AGGTTTCGCT TTCCCACTGA CTTGTCTCCA ATCAGTTTAC AATGACAACT
AfMNPVK2                        CAACATATAC AGAATTAGTC TGGCTGATAA GGGCGGCGGC TGCCCAATCA TGAACATTCA CAGCGAATAT ACTAATTCGT
Aglais_urticae_NPVA7            CAACGAATAC CGCATCAGTC TGGCCAAGAA AGGCGGCGGT TGCCCGGTGA TGAACCTTCA CTCAGAGTAC ACAAACTCAT
Agraulis_vanillare_NPV779       CAACGAGTAC CGCATCAGYC TGGCCAAAAA GGGMGGCGGC TGTCCCATMA TGAACATTCA CAGCGAGTAC ACCAAYTCGT
Agrotis_segetum_NPVi6           CAACGAGTAC CGCATTAGTC TCGCCAAGAA GGGCGGTGGC TGCCCCGTAA TGAATCTGCA CTCTGAGTAT TCCAACTCCT
Agrotis_segetum_NPVK4           CAACGAGTGC CGCATTAGTC TCACCAAGAA GGGCGGAGGC TGCCCCGTAA TGAATCTGCA CTCTGAGTAT TCCAACTCCT
Ammalo_insulata_NPV175          CAACGAATAC AGAATTAGCC TTGCCAAAAA AGGCGGTGGC TGCCCCATCA TGAACATTCA CAGCGAGTAC ACCAACTCGT
Amonis_sabulifera_NPV14         TAACGAATAC AGAATCAGCC TGGCCAAGAG GGGAGGCGGT TGCCCCGTCA TGAACCTTCA CGCCGAGAAC ACCAACTCAT
Amsacta_albistriga_NPV          CAACGAATAC CGTATTAGTT TAGCCAAAAA AGGTGGCGGT TGTCCCATTA TGAACATCCA CCGAGAATAT ACCAACTCTT
Antherea_pernyi_NPV             CAACGAGTAC AGAATTAGCC TCGCCAAGAA AGGCGGCGGC TGCCCCATCA TGAACATTCA CAGCGAGTAC ACCAACTCGT
Antherea_polyphemus_NPV30       CAACGAGTAC AGAATTAGTC TGGCTAAGAA GGGCGGCGGC TGCCCAATCA TGAACATCCA CAGCGAGTAC ACCAACTCGT
Anthophila_fabricana_GV790      CAATGTCGAA CGTATTAACC TCACTAAAAA GGGATTCGCA TTCCCTCTGA CCTGTCTGCA GTCCGTGTAC AATGACAACT
Anticarsia_gemnitalis_NPV       CAACGAGTAC AGAATTAGCC TAGCCAAGAA AGGGGGTGGT TGCCCAATCA TGAACATCCA CAGCGAGTAT ACCAACTCGT
Aporia_crataegi_NPV106          CAACGAGTAC AGAATTAGTT TAGCTAAAAA GGGCGGCGGA TGTCCCATTA TGAACATTCA CAGCGAGTAC ACCAACTCGT
Archips_cerasivora_NPV          CAACGAGTAC CGCATCAGCC TTGCAAAGAA AGGAGGCGGC TGCCCCATCA TGAACATCCA CGCTGAGTAC ACCAACTCGT
Aroa_discalis_NPV63             CAACGAATAC AGAATCAGTT TGGCTAAACG CGGAGGAGGA TGCCCCGTCA TGAACCTACA CGCCGAGTAC ACCAACTCGT
Artogeia_napi_GV242             CAATATTGAA CGCATCAATC TG---AAGAA AGGTTTCGCT TTCCCGCTTA TGTGCCTTCA ATCCGTGTAC AATGACAACT
Attacus_ricini_NPV              CAACGAGTAC AGAATTAGCC TGCCCAAGAA AGGCGGCGGC TGCCCCATCA TGAACATTCA CAGCGAGTAC ACCAACTCGT
BmNPV                           CAACGAATAC AGAATTAGTC TGGCTAAAAA GGGCGGCGGC TGCCCAATCA TGAACATCCA CAGCGAGTAC ACCAACTCGT
Bombyx_mori_NPV460              CAACGAATAC AGAATTAGTC TGGCTAAAAA GGGCGGCGGC TGCCCAATCA TGAACATCCA CAGCGAGTAC ACCAACTCGT
Buzura_edwardsii_NPV505         CAACGAATAC CGCATCAGCC TGGCCAAGAA GGGCGGCGGC TGCCCTGTCA TGAACCTTCA CTCTGAGTAC ACGCACTCCT
Buzura_Supressaria_NPV          CAACGAGTAT CGCATTAGTT TAGCCAAAAG AGGCGGTGGT TGCCCGGTAA TGAATCTGCA TTCTGAATAC ACCAACTCGT
Cacoecia_murinana_GV37          CAACATGGAA CGCATCAACT TGTCCAAGAA AGGGTTCGCA TTCCCGCTCA CGTGTCTTCA GTCCGTTTAC AATGACAACT
Catopsila_pomona_NPV78          CAACGAGTAC AGAATTAGTT TGGCCAAAAA AGGAGGCGGC TGCCCCATAA TGAACATTCA CAGCGAGTAC ACCAACTCGT
Cerapteryx_gramini_NPVA5        CAACGAGTAC CGCATCAGYC TGGCCAAGAA GGGMGGCGGC TGCCCCGTGA TGAACYTGCA CGCCGAGTAC ACCAACTCGT
ChfuNPV                         CAACGAATAC CGCATCAGCC TGGCGAAAAA AGCCGCCGGC TGTCCCATTA TGAACATCCA CCCCGAATAC ACCAATTCGT
ChfuNPVDEF                      CAACGAGTAC AGAATTAGCC TAGCCAAGAA AGGCGGTGGC TGCCCCATCA TGAACATCCA CAGCGAGTAC ACCAACTCGT
Choristoneura_furamifera_GV     CAACATGGAA CGCATCAACT TGTCCAAGAA AGGATTCGCA TTCCCGCTCA CGTGTCTTCA ATCCGTCTAC AATGACAACT
Choristoneura_rosaceana_NPV     CAACGAATAC CGCATCAGCC TGGCGAAAAA AGGCGGCGGC TGTCCCATTA TGAACATCCA CGCCGAATAC ACCAATTCGT
Cirphis_unipuncta_NPV172        CAACGAATAC CGAGTCAGCT TGGCCAAGAG AGGCGGTGGC TGCCCCATGA TGAATCTGCA CTCTGAGTAC ACCAACTCTT
CpGV                            CAACATTGAG CGCATCAACC TCTCCAAAAA GGGCTTTGCG TTCCCGCTCA CTTGTCTTCA GTCTGTCTAC AACGACAACT
Cryptophlebia_leucotreta_GV     CAATATCGAG CGTATCAATC TTTCCAAGAA GGGTTTCGCT TTCCCACTCA CATGCCTACA GTCCGTCTAC AATGACAACT
Cynosarga_chrysolopa_NPV302     CAACGAATAC AGAATCAGCT TGGCCAAAAA AGGTGGCGGC TGCCCAATTA TGAACATTCA CAGCGAGTAC ACCAACTCGT
Darna_trima_GV545               CAACATTGAA CGTATCAATT TGTCAAAGAA AGGTTTTGCT TTCCCTCTCA CGTGTTTGCA ATCAGTTTAC AATGACAATT
Diacrisia_obliqua_NPV798        CAACGAGTAC CGCATCAGCT TGGCCAAACG AGGCGGCGGC TGCCCCATAA TGAACATTCA CAGCGAATAC ACCAACTCAT
Diatrea_saccharalis_NPVK6       CAATATCGAG CGAATCAATT TGACCAAGAA AGGATTTGCG TTCCCATTGA CGTGTCTACA GTCTGTGTAC AATGAAAACT
Ectropis_obliqua_NPV            CAACGAATAT CGCATCAGTC TGGCTAAACG CGGAGGCGGT TGCCCCGTTA TGAACCTGCA CGCCGAGTAC ACCAACTCTT
Ephestia_cautella_NPVC5         CAATGAATAC AGGATCAGTT TGGCCAAGAA GGGAGCTTGC TGTCCTATTA TGAATATTCA TGGAGAATAC ACTAATTCCT
EppoMNPV                        CAACGAATAC AGAATTAGTT TAGCCAAAAA GGGCGGCGGC TGCCCCATTA TGAACATTCA CAGCGAGTAC ACTAACTCGT
Erannis_defolariaNPV174         CAACGAATAT CGCATTAGCT TAGCCAAGAA AGGCGGCGGA TGCCCAGTCA TGAACCTGCA CGCCGAGTAC ACCAACTCGT
Euplasia_luciparaGV248          CAACATCGAG CGCATCAACC TCTCCAAGCG CGGAGTCGCG TTCCCTCTCA CATGCCTACA GTCTGTGTAC AATGACAACT
Euproctis_Chrysorrohea_NPVA10   CAACGAATAC AGAATTAGTT TAGCCAAAAA AGGTGGCGGC TCTCCAATTA TGAACATTCA CAACGAATAC ACCAATTCGT
Euproctis_similis_NPV768        CAACGAATAC AGAATTAGTT TAGCCAAAAA AGGTGGCGGC TCTCCAATTA TGAACATTCA CAACGAATAC ACCAATTCGT
Harrisina_brillians_GV          CAATATTGAA CGTATAAATT TGACAAAAAA AGGCTTTGCT TTCCCGTTAA CTTGTCTTCA ATCTGTGTAC AATGATAACT
Harrisina_brillians_GVM2        CAATATTGAA CGTATAAATT TGACAAAAAA AGGCTTTGCT TTCCCGTTAA CTTGTCTTCA ATCTGTGTAC AATGATAACT
HaSNPV                          CAACGAGTAC AGAATTAGTT TAGCCAAAAA ATACGGCGGT TGTCCCGTTA TGAACTTGCA CGCTGAATAC ACTAATTCCT
Heliconius_erato_NPV789         CAACGAGTAC CGCATCAGCC TGGCCAAAAA GGGCGGCGGT TGTCCCATCA TGAACATTCA CAACGAGTAC ACCAATTCGT
Homona_coffearia_GV744          CAACATGGAA CGCATCAACT TGTCCAAAAA AGGATTTGCA TTCCCGCTTA CGTGTCTTCA GTCGGTGTAC AATGACAACT
Homona_coffearia_NPV745         CAACATGGAA CGCATCAACT TGTCCAAAAA AGGATTTGCA TTCCCGCTTA CGTGTCTTCA GTCGGTGTAC AATGACAACT
Hyphantria_cunea_GV228          CAACACGGAA CGCATCAACT TGACTCGAAA GGGATGCGCG TTCCCACTAA CATGCTTACA GGCAGCTTAC AACGATTGTT
Hyphantria_cunea_NPV            CAACGAGTAC CGCATCAGCT TGGCCAAACG AGGCGGCGGC TGCCCCATAA TGAACATTCA CAGCGAATAC ACCAACTCAT
Hyphantria_cunea_NPVN3          CAACGAGTAC CGCATCAGCT TGGCCAAACG AGGCGGCGGC TGCCCCATAA TGAACATTCA CAGCGAATAC ACCAACTCAT
HzSNPV_genome                   CAACGAGTAC AGAATTAGTT TAGCCAAAAA ATACGGCGGT TGTCCCGTTA TGAATTTGCA CGCTGAATAC ACTAATTCCT
Idaea_virgulata_NPV402          TAACGAGTAC AGAATCAGCC TAGCCAAGAA AGGCGGCGGC TGTCCCATTA TGAATATCCA CAGCGAATAC ACTAACTCGT
Junonia_coenia_GV19             CAACATTGAG CGCATTAATC TATCGAAGAA AGGTTACGCG TATCCTTTGA CGTGTCTGCA GTCGGTGTAC AATGAAAATT
Kotochalia_junodi_NPV205        TAACGAATAT CGCATCAGTC TAGCTAAAAG AGGAGGCGGT TGTCCCATCA TGAATCTGCA TGCCGAGTAC ACCAATTCCT
Kotochalia_junodi_NPV32         TAACGAATAT CGCATCAGTC TAGCTAAAAG AGGAGGCGGT TGTCCCATCA TGAATCTGCA TGCCGAGTAC ACCAATTCCT
Lacanobia_orecalcea_GVA6        CAACATCGAG CGCATCAACC TGTCCCGCAA GGGAGTTGCG TTCCCTCTAA CCTGCCTGCA GTCCGTGTAC AACGACAACT
Lasiocampa_quercus_NPV202       CAACGAGTAT CGCATCAGTC TTGGCAAACG CTACAACGGT TGTCCCGTCA TGAACCTGCA CTCGGAGTAC ACCAACTCTT
LdMNPV                          CAACGAGTAC AGGATTAGCC TGGCTAAGCG AGGCGGCGGC TGTCCCATCA GGAACTTGCA CAGCGCGTAC ACGACCTCGT
Lonomia_obliqua_NPV             CAACGAATAT AGAATCAGTT TAGCCAAAAA AGGCGGCGGT TGCCCAATCA TGAACATTCA CAGCGAATAC ACCAATTCTT
Lymantria_monacha_NPVB4         CAACGAGTAC CGCATCAGTC TGGCCAAGAG AGGCGGTGGC TGCCCCATCA GAAACCTTCA CAGCGCCTAC ACCACGTCGT
Lymantria_nanayi_NPV750         CAACGAGTAC AGAATCAGCC TGGCCAAGCG AGGCGGCGGC TGTCCCATTA GAAACTTGCA CAGCGCGTAC ACGACCTCGT
Lymantria_obfuscata_NPV193      CAACGAGTAC AGGATCAGCC TGGCTAAGCA AGGCGGCGGC TGTCCCATCA TGAACTTGCA CAGCGCGTAC ACGACCTCGT
Mahasena_corbetti_GV699         CAACGAGTAC AGAATCAGTC TGGCCAAGAA AGGCGGCGGT TGCCCTATAT TGAACCTTCA TCCAGAGTAC ACCAACTCTT
Mahasena_corbetti_NPV595        TAATGAGTAT AGAATCAGTT TAGCTAAACG AGGTGGTGGT TGCCCTATCA TGAACCTTCA CGCCGAATAC ACTAATTCAT
Mahasena_corbetti_NPV692        TAATGAGTAT AGAATCAGTT TAGCTAAACG AGGTGGTGGT TGCCCTATCA TGAACCTTCA CGCCGAATAC ACTAATTCAT
Malacosoma_neustria_NPV         CAACGAGTAC CGCATCAGTT TGGGCAAACG ATACAACGGA TGCCCAGTTA TGAACTTGCA TTCAGAGTAC ACCAATTCCT
Malacosoma_neustria_NPVk12      CAACGAGTAC CGCATTAGTT TGGGTAAACG GTACAACGGA TGCCCAGTAA TGAACTTGCA TTCAGAGTAC ACCAATTCCT
Malacosoma_pluviale_NPVk3       CAACGAGTAC CGCATTAGTT TGGGTAAACG GTACAACAGA TGCCCATTAA TGAACTTGCA TTCACAGTAC ACCAATTTCT
Malocosoma_disstria_NPV         CAACGAGTAC CGCATCAGTC TGGGCAAACG ATACAACGGA TGCCCGGTTA TGAACTTGCA TTCGGAATAC ACCAACTCGT
Mamestra_brassicae_NPV          CAACGAATAC CGCGTCAGCT TAGCCAAGCG TGGCGGTGGC TGCCCCGTGA TGAATCTGCA CTCTGAATAC ACCAACTCTT
Mamestra_brassicae_NPVK9        CAACGAATAC CGCGTGAGCT TAACCAAGCG TGGCGGGGGC TGCCCCGTGA TGAATCTGCA CTCTGAATAC ACCAACTCTT
Mamestra_configurata_NPVA       CAACGAATAC CGTGTCAGCT TGGCCAAGCG AGGCGGTGGC TGCCCCGTGA TGAACCTGCA CTCTGAGTAC ACCAACTCTT
Mamestra_configurata_NPVB       CAACGAATAC CGCGTCAGCT TAGCCAAGCG TGGCGGTGGC TGCCCCGTGA TGAATCTGCA CTCTGAATAC ACCAACTCTT
Melanchra_persicariaGV26        CAACATTGAG CGCATCAACC TGTCCCGCAA GGGTGTCGCT TTCCCTCTAA CCTGTCTGCA GTCCGTGTAC AACGACAACT
Mynthimna_pseudolatia_NPVA1     CAACGAGTAC CGCGTCAGCC TGGCTAAACG CGGCGGCGGT TGCCCCATCA TGAACCTTCA CTCTGAGTAC ACCAACTCGT
'Mynthimna separata (Lese)NPV'  CAACGAATAC CGCGTCAGCT TAGCCAAGCG TGGCGGTGGC TGCCCCGTGA TGAATCTGCA CTCTGAGTAC ACCAACTCTT
Mynthimna_separata_ChinaNPVA3   CAACGAGTAC CGCATCAGTC TGGCCAAGAA GGGAGGCGGC TGTCCCGTGA TGAACTTGCA CGCCGAGTAC ACCACCTCGT
Mynthimna_separata_IndiaNPVA4   CAACGAGTAC CGCATCAGTC TGGCCAAGAA GGGAGGCGGC TGTCCCGTGA TGAACTTGCA CGCCGAGTAC ACCACCTCGT
Mynthimna_separata_Japan_NPVA2  CAACGAATAC CGCGTCAGCT TAGCCAAGCG TGGCGGTGGC TGCCCCGTGA TGAATCTGCA CTCTGAGTAC ACCAACTCTT
Natada_nararia_GV254            CAACATCGAA CGTATTAACT TGTCTAAAAA AGGATTCGCA TTCCCATTGA CCTGTCTCCA ATCCGTATAC AATGATAACT
Nymphalis_io_NPV353             CAACGAATAC CGCATCAGTC TGGCCAAGAA AGGCGGCGGT TGCCCGGTGA TGAACCTTCA CTCAGAGTAC ACAAACTCAT
Operoptera_brumata_NPV291       CAACGAATAC AGAATTAGTC TGGCCAAGAA GGGCGGCGGT TGCCCGCTAT TGAACCTTCA TCAAGAGTAC ACCAACTCTT
Operoptera_brumata_NPVi4        CAACGAATAC AGAATTAGTC TGgCCAAGAA GGCGGCAGGT TGCCCGCTAT TGAACCTTCA TCAAGAGTAC ACCAACTCTT
Operoptera_brumata_NPVi5        CAACGAATAC AGAATTAGTC TGGCCAAGAA GGGCGGCGGT TGCCCGCTAT TGAACCTTCA TCAAGAGTAC ACCAACTCTT
OpMNPV                          CAACGAGTAC CGCATCAGCC TGGCCAAAAA AGGCGGCGGC TGCCCTATCA TGAACATTCA CGCCGAATAC ACCAACTCGT
OpSNPV                          CAACGAGTAC AGAATTAGTC TGGCTAAACG CGGCGGCGGT TGCCCCGTGA TGAACTTACA CGCCGAGTAC ACCAACTCTT
Orgya_antiqua_NPVA9             CAACGAGTAC AGAATTAGTC TGGCTAAACG CGGCGGCGGT TGCCCCGTGA TGAACTTACA CGCCGAGTAC ACCAACTCTT
Orgya_leucostigma_NPVB6         CAACGAGTAC AGAATCAGTT TGGCCAAACG AGGAGGCGGA TGTCCCGTCA TGAATTTACA CGCCGAGTAC ACAAACTCGT
Orgya_mixta_NPV67               CAACGAATAC AGAATCAGTT TGGCTAAACG CGGAGGAGGA TGCCCCGTCA TGAACCTACA CGCCGAGTAC ACCAACTCGT
Orgya_pseudotsugattaNPVB5       CAACGAGTAC AGCATCAGCC TGGCCAAAAA AGGAGGCGGA TGCCCTATCA TGAACTTTCA CTCCGAGTAC ACAAACTCGT
Orgyia_anartoides_NPV           CAACGAATAC AGAATAAGTC TGGCGAAACG AGGCGGTGGT TGCCCCGTAA TGAACCTGCA CGCAGAGTAC ACAAACTCTT
Pachypasa_papyri_NPV436         TAATGAATAT CGTATCAGTC TGGGTAAACG CTATAACGGT TGTCCGGTAA TGAATCTCCA TTCGGAGTAC ACCAACTCTT
Panolis_flammea_MNPV            CAACGAATAC CGTGTCAGCT TGGCCAAGCG AGGCGGTGGC TGCCCCGTGA TGAACCTGCA CTCTGAGTAC ACCAACTCTT
Panolis_flammea_NPVi3           CAACGAATAC CGTGTCAGCT TGGCCAAGCG AGGCGGTGGC TGCCCCGTGA TGAACCTGCA CTCTGAGTAC ACCAACTCTT
Perina_nuda_NPV                 CAACGAGTAC CGCATCAGCC TGGCCAAAAA AGGCGGCGGC TGCCCCATCA TGAACATTCA CGCCGAATAC ACCAATTCCT
PhopGV                          CAATGTCGAG CGCATCTGTC TATCCAAACA GGGTGTTGCT TTCCCGCTCA CTTGCCTTCA ATCCGTGTAC AATGACAACT
```

```
Phthorimaea_operculella_GVC3     CAATGTCGAG CGCATCTGTC TATCCAAACA GGGTGTTGCT TTCCCGCTCA CTTGCCTTCA ATCCGTGTAC AATGACAACT
Phthorimaea_operculella_GVC4     CAATGTCGAG CGCATCTGTC TATCCAAACA GGGTGTTGCT TTCCCGCTCA CTTGCCTTCA ATCCGTGTAC AATGACAACT
Pieris_brassicae_GV              CAATATTGAA CGCATCAATC TG---AAGAA AGGTTTCGCT TTCCCGCTTA TGTGCCTTCA ATCCGTGTAC AATGACAACT
Pieris_brassicae_GV384           CAATATTGAA CGCATCAATC TG---AAGAA AGGTTTCGCT TTCCCGCTTA TGTGCCTTCA ATCCGTGTAC AATGACAACT
Pieris_rapaeGV95                 CAATATTGAA CGCATCAATC TG---AAGAA AGGTTTCGCT TTCCCGCTTA TGTGCCTTCA ATCCGTGTAC AATGACAACT
Plodia_interpunctella_GVB3       TAACGTCGAG CGCATCAACC TCACCAAGAA AGGATTCGCG TTYCCTCTGA CCAGTCTGCA GTCCGTCTAC AATGACAACT
Plusia_gammaNPV287               CAACGAGTAC CGCATCAGCC TGGCTAAGAA GGGCGGCGGC TGCCCAATAA TGAACCTTCA CTCTGAGTAC ACCAACTCGT
Plusia_orichalcea_NPV            CAACGAATAC CGCATCAGCT TGGCTAAGAA AGGCGGTGGC TGCCCAATCA TGAACCTTCA TTCCGAGTAC ACCAACTCGT
'Polygonia c-album NPV380'       CAACGAATAC CGCATCAGTC TGGCCAAGAA AGGCGGCGGT TGCCCGGTGA TGAACCTTCA CTCAGAGTAC ACAAACTCAT
PxGV                             CAACGTTGAG CGCATCAGTC TGTCCCGACA GGGTGTGGCG TTCCCCCTCA CTTGCCTGCA GGCCGTTTAC AACGAGAGCT
RaouNPV                          CAACGAGTAC AGAATTAGTC TGGCTAAGAA GGGCGGCGGC TGCCCAATCA TGAACATTCA CAGCGAATAT ACTAATTCGT
Samia_ricini_NPV28               CAACGAGTAC CGCATCAGCC TGGCTAAGAA GGGCGGCGGC TGCCCAATAA TGAACCTTCA CTCTGAGTAC ACCAACTCGT
Shpinx_ligustri_NPV334           CAACGAGTAC AGAATTAGTC TGGCTAAGAA GGGCGGCGGC TGCCCAATCA TGAACATCCA CAGCGAGTAC ACCAACTCGT
Smerinthus_ocellata_NPV819       CAATGAATAC CGCATCAGTC TCGCCAAGAA AGGTGGCGGC TGTCCCGTGA TGAACCTGCA CGCCGAATAC ACCACTTCGT
Sphinx_ligustri_NPV362           CAACGAGTAC CGCATCAGCC TGGCTAAGAA GGGCGGCGGC TGCCCAATAA TGAACCTTCA CTCTGAGTAC ACCAACTCGT
Spilosoma_lutea_GV126            CAACACGGAG CGCATCAATC TTACCCGCAA GGGTTGCGCC TTCCCCCTTA CGTGCTTACA GGCCGTATAC AACGATTGTT
Splisoma_imparilis_NPVN2         CAACGAGTAC CGCATCAGCT TGGCCAAACG AGGCGGCGGC TGCCCCATAA TGAACATTCA CAGCGAATAC ACCAACTCAT
Spodoptera__exempta_NPVK11       CAACGAGTAC CGCATCAGCC TGGCCAAGAA GGGCGGCGGT TGCCCCGTGA TGAACCTGCA TTCCGAGTAC ACCAACAGCT
Spodoptera_exigua_MNPV           CAACGAATAC CGCATCAGTT TAGCCAAAAA GGGCGGCGGT TGCCCCGTCA TGAATCTCCA CTCCGAGTAC ACCAACTCGT
Spodoptera_frugiperda_NPV        CAACGAATAC CGCATCAGCC TGGCCAAGAA GGGCGGCGGC TGCCCTGTCA TGAACCTTCA CTCTGAGTAC ACGCACTCCT
Spodoptera_frugiperda_NPVK10     CAACGAATAC CGCATCAGCC TGGCCAAGAA GGGCGGCGGC TGCCCTGTCA TGAACCTTCA CTCTGAGTAC ACGCACTCCT
Spodoptera_litoralis_NPV         CAATGAATAC CGCATCAGTC TCGCCAAGAA AGGTGGCGGC TGTCCCGTGA TGAACCTGCA CGCCGAATAC ACCACTTCGT
Spodoptera_littoralis_GV66       CAACCGCGAG CGTATCAATT TGAGTAAGAA GGGTGAAGCC TTCCCTCTGA CCTGTCTCCA ATCGATTTAT AATGAGAACT
Spodoptera_littoralis_NPV440     CAATGAATAC CGCATCAGTC TCGCCAAGAA AGGTGGCGGC TGTCCCGTGA TGAACCTGCA CGCCGAATAC ACCACTTTGT
Spodoptera_litura_GV766          CAACCGCGAG CGTATCAATT TGAGTAAGAA GGGTGAAGCC TTCCCTCTGA CCTGTCTCCA ATCGATTTAT AATGAGAACT
Spodoptera_litura_NPVg           CAACGAGTAC CGCATCAGCT TGGCCAAAAA GGGCGGCGGC TGCCCGGTCA TGAATCTACA CTCTGAGTAC ACGCACTCTT
Spodoptera_litura_NPV2           TAACGAATAC CGCATCAGTC TGGCTAAGAA GGGCGGCGGT TGTCCCGTGA TGAACCTGCA CGCCGAATAC ACCACGTCGT
Spodoptera_sunia_NPVK8           CAACGAGTAC AGAATCAGTC TGGCTAAAAA GGGCGGCGGC TGCCCCATCA TGAACCTCCA CTCTGAATAC ACCAACTCGT
Spodoptera_sunia_NPVK7           CAACGAATaC CGCATcAGTC TCACCAAGAA AGGTGGAGGC TGCCCCGTCA TGAACCTTCA TTCCgAGTAC ACCAACTCGT
Thysanoplusia_orichalcea_NPV     CAACGAGTAC AGAATTAGTT TAGCCAAAAA GGGCGGCGGC TGCCCGATCA TGAACATACA CAGCGAGTAC ACCAACTCGT
Thysanoplusia_orichalcea_NPV2    CAACGAaTAC cGcATcAGcT TgGCtAAgAA GGGCGGtGGC TGCCCaATCA TGAACcttCA ttcCGAGTAC ACCAACTCGT
Trabala_vishnou_NPV195           CAACGAGTAC AGAATCAGTT TAGCCAAAAA AGGTGGCGGC TGCCCAATAA TGAATCTGCA CGCCGAATAC ACGAATTCAT
Trichoplusia_ni_GV               CAACCGTGAA CGCATCAACC TTTCTAAGAG AGGTCTTGCC TTCCCTCTGA CTTGTCTTCA GTCTATCTAC AACGAAAACT
Trichoplusia_ni_MNPVC2           CAACGAGTAC CGCATCAGCC TGGCTAAGAA GGGCGGCGGC TGCCCAATAA TGAACCTTCA CTCTGAGTAC ACCAACTCGT
Trichoplusia_ni_SNPV             CAACGAATAC AGAATTAGTC TGGCCAAGAA AGGCGGTGGC TGTCCCATCA TGAACCTTCA CTCTGAGTAC ACCAACTCGT
Trichoplusia_orichalceaNPVB9     CAACGAGTAC CGCATCAGCC TGGCCAAGAA GGGAGGCGGC TGCCCGATCA TGAACTTACA CTCTGAGTAC ACCAACTCGT
Utetheisa_pulchella_NPV142       CAACGAATAC AGAATTAGTC TGGCCAAGAA GGGCGGCGGC TGCCCCATAT TGAACCTTCA TCAAGAGTAC ACCAACTCTT
Vanessa_alatanta_NPV54           CAACGAATAC CGCATCAGTC TGGCCAAGAA AGGCGGCGGT TGCCCGGTGA TGAACCTTCA CTCAGAGTAC ACAAACTCAT
Vanessa_cardui_NPV367            CAACGAATAC CGCATCAGTC TGGCCAAGAA AGGCGGGGGT TGCCCGGTGA TGAACCTTCA CTCAGAGTAC ACAAACTCAT
Wiseana_cervinata_GV342          TAACATAATA AGGATAAATC TTGTCAAGAA AGGATTCGGC CTTCCCCTCA ATTATCTTCA GTCCGTATAC AATGACTACT
Wiseana_cervinata_NPV344         CAACGAGTAC AGGATCAGCC TCGCCAAGAG GGGAGGCGGC TGCCCCATCA TGAATCTTCA CTTGGAGTAC ACCAACTCTT
Wiseana_signata_NPV              CAACGAGTAC AGGATCAGCC TCGCCAAGAG GGGAGGCGGC TGCCCCATCA TGAATCTTCA CTTGGAGTAC ACCAACTCTT
XcGV                             CAACCGTGAA CGCATCAACC TTTCCAAGAG AGGTCTCGCC TTCCCTCTGA CGTGCCTGCA GTCGATCTAC AACGAAAACT


Abraxasgrossulariata_NPV112      TTGAGTCGTT TGTAAACCGT GTCATCTGGG AAAACTTCTA CAAGCCCATC GTGTACATTG GCACGGATTC GGGCGAAGAA
Achaea_faber_NPV64               TCGAAGAGTT CATCAACCGT GTCATTTGGG AGAACTTCTA CAAACCCATC GTGTACATCG GTACCGATTC CGGAGAAGAG
AchaeajanataGV835                TTGAAGAGTT CTTTGACAAC GTCATCTGGC CCTACTTCCA CCGGCCACTC GTCTACGTCG GTACAACCTC GGCCGAGATC
AcMNPV                           TCGAACAGTT CATCGATCGT GTCATCTGGG AGAACTTCTA CAAGCCCATC GTTTACATCG GTACCGACTC TGCTGAAGAG
AcMNPVK1                         TCGAACAGTT CATCGATCGT GTCATCTGGG AGAACTTCTA CAAGCCCATC GTTTACATCG GTACCGACTC TGCTGAAGAG
Actias_selene_NPV47              TTGAATCGTT TGTAAACCGG gTAATCTGGG AGAACTTTTA CAAGCCCATT GAGTACATCG GCACGGACTC GGGTGAGGAG
Adoxophyes_orana_GV              TTGAAAGGTT CTTTGATGAT GTGCTGTGGC CCTATTTCCA TCGCCCACTA GTCTATATCG GTACCACGTC TGCCGAAGTC
AfMNPVK2                         TCGAGTCGTT TGTGAGCCGC GTCATATGGG AGAATTTCTA CAAACCCATC GTTTACATTG GCACAGACTC TGGCGAAGAA
Aglais_urticae_NPVA7             TCGAAGAGTT CATTTCTAAA GTAATTTGGG AGAACTTCTA CAAACCAATC GTGTACCTCG GTACCGATTC CGCTGAAGAA
Agraulis_vanillare_NPV779        TTGAATCKTT CGTGRASCGC GTMATTTGGG AAAACTTTTA CAARCCCATY GTGGACATTG GCACCGATTC TGGYGAAGAG
Agrotis_segetum_NPVi6            TTGAAGAGTT TATCAACCGC GTCATATGGG AAAACTTCTA CAAGCCCATC GTGTATATCG GTACAGACTC TGCTGAGGAA
Agrotis_segetum_NPVK4            TTGAAGAGTT TATCAACCGC GTCATATGGG AAAACTTCTA CAAGCCCATC GTGTATATCG GTACAGACTC TGCTGAGGAA
Ammalo_insulata_NPV175           TTGAATCGTT TGTCAACCGC GTAATCTGGG AGAACTTTTA CAAGCCCATC GTGTACATTG GCACAGACTC GGGCGAGGAG
Amonis_sabulifera_NPV14          TCGAACAGTT CATCGCTCGC GTGATCTGGG AGAACTTCTA CAAGCCCATC GTGTACGTGG GCACCGATTC CGCCGAAGAG
Amsacta_albistriga_NPV           TTGAGTCTTT TGTCAACCGT GTTATTTGGG AAAACTTTTA CAAGCCTATC GTGGACATTG GCACGGATTC GGGTGAAGAG
Antherea_pernyi_NPV              TTGAATCGTT TGTAAACCGC GTAATCTGGG AGAACTTTTA CAAGCCCATT GTGTACATCG GCACGGACTC GGGTGAGGAG
Antherea_polyphemus_NPV30        TCGAGTCGTT TGTGAACCGC GTCATATGGG AGAACTTCTA CAAACCCATC GTTTACATCG GCACAGACTC TGGCGAAGAA
Anthophila_fabricana_GV790       TTGAATCGTT CTTTGACGAT ATTCTGTGGC CATATTTCCA CAGACCTTTG GTCTATGTCG GTACAACATC GGGAGAAATT
Anticarsia_gemnitalis_NPV        TTGAGTCCTT TGTCAACCGC GTAATCTGGG AAAACTTTTA CAAACCCATT GTGTACATTG GCACTGATTC GGGCGAAGAA
Aporia_crataegi_NPV106           TTGAATCTTT CGTCAACCGT GTAATCTGGG AGAACTTTTA CAAACCAATT GTGTACGTAG GAACCGATTC AGCCGAAGAA
Archips_cerasivora_NPV           TCGAATCCTT TGTGAACCGC GTCATCTGGG AAAACTTTTA CAAGCCCATC GTGTACATTG GCACAGACTC CGGCGAAGAG
Aroa_discalis_NPV63              TTGAAGAGTT TATCAACCGC GTTATCTGGG AGAACTTCTA CAAGCCCATC GTGTACGTGG GAACCGATTC GGCCGAGGAA
Artogeia_napi_GV242              TTGAAACTTT CTTTGAAGAT GTACTATGGC CATACTTCCA CAGACCACTC GTCTACATCG GTACCACCTC TAGCGAAACC
Attacus_ricini_NPV               TTGAATCGTT TGTAAACCGC GTAATCTGGG AGAACTTTTA CAAGCCCATT GTGTACATCG GCACGGACTC GGGTGAGGAG
BmNPV                            TCGAGTCGTT TGTGAACCGC GTCATATGGG AGAACTTCTA CAAACCCATC GTTTACATCG GCACAGACTC TGCCGAAGAA
Bombyx_mori_NPV460               TCGAGTCGTT TGTGAACCGC GTCATATGGG AGAACTTCTA CAAACCCATC GTTTACATCG GCACAGACTC TGCCGAAGAA
Buzura_edwardsii_NPV505          TCGAAGAGTT CATCGACCGT GTCATCTGGG AGAACTTCTA CAAACCCATC GTGTACGTAG GAACCGACTC TGGTGAGGAA
Buzura_Supressaria_NPV           TCGAGGAATT CATCAACAGA GTAATTTGGG AAAACTTTTA CAAACCTATT GTGTACGTAG GTACCGATTC GGCAGAGGAA
Cacoecia_murinana_GV37           TCGAGAGATT CTTCGACGAC ATTCTTTGGC CGTATTTCCA CCGTCCTCTA GTGTATGTGG GAACCACTTC CGCCGAAATC
Catopsila_pomona_NPV78           TTGAATCGTT CGTGAACCGT GTCATTTGGG AGAACTTTTA CAAGCCCATT GTTTACATTG GTACCGATTC TGCTGAAGAA
Cerapteryx_gramini_NPVA5         TYGAAAAAAT CATCAACCGA GTCATATGGG AMAACTTCTA CAAGCCCATC GTGTACATYG GAACCGAYTC TGGCGAAGAA
ChfuNPV                          TCGAATCCTT TGTGAACCGC GTCATCTGGG AAAACTTTTA CAAACCCATC GTGTACATTG GCACAGATTC AGGCGAGGAA
ChfuNPVDEF                       TTGAGTCCTT TGTCAACCGC GTAATCTGGG AAAACTTTTA CAAACCCATC GTGTACATTG GCACTGACTC GGGCGAGGAA
Choristoneura_furamifera_GV      TTGAGAGATT CTTCGACGAC ATTCTTTGGC CGTATTTCCA CCGTCCTTTA GTATATGTGG GAACCACTTC CGCCGAAATC
Choristoneura_rosaceana_NPV      TCGAATCCTT TGTGAACCGC GTCATCTGGG AAAACTTTTA CAAACCCATC GTGTACATTG GCACAGATTC AGGCGAGGAA
Cirphis_unipuncta_NPV172         TCGAAGAGTT TATCAACCGT GTCATCTGGG AGAACTTCTA CAAGCCCATC GTGTACGTTG GTACCGATTC GGCTGAGGAA
CpGV                             TTGAACGCTT CTTTGACGAT GTGCTGTGGC CCTACTTCTA CCGACCACTG GTGTACGTTG GCACCACCTC CGCCGAAATT
Cryptophlebia_leucotreta_GV      TTGAGAATTT CTTTGATGAC GTTCTGTGGC CGTATTTCCA CCGTCCCCTG GTGTATGTAG GTACTACGTC TGCCGAAATT
Cynosarga_chrysolopa_NPV302      TTGAATCGTT TGTCAACCGC GTGATTTGGG AAAACTTTTA CAAACCCATC GTGTACATCG GTACCGACTC TGGCGAAGAA
Darna_trima_GV545                TTGAAAGATT CTTTGACGAC GTTCTGTGGC CTTATTTCCA CAGACCTTTG GTTTATGTTG GAACAACTTC TGCCGAAATC
Diacrisia_obliqua_NPV798         TTGAACAGTT CGTTAACCGC GTAATTTGGG AGAACTTTTA CAAACCCATC GTGTACATTG GCACAGACTC TGGCGAGGAA
Diatrea_saccharalis_NPVK6        TTGAATGTTT CTTCGACAAT GTTCTCTGGC CATATTTCTA TCGTCCTCTA GTTTATGTGG GAACTACATC TGCCGAAATA
Ectropis_obliqua_NPV             TCGAGGAATT CATCAACCGC GTGATTTGGG AGAATTTCTA CAAGCCCATC GTGTATGTGG GCACCGATTC CGCCGAAGAG
Ephestia_cautella_NPVC5          TTGAACACTT TATTAACAAG GTGATCTGGG AGAACTTCTA CAAACCCATT GTGTATATTG GTACAGATTC TAGTGATGAG
EppoMNPV                         TTGAATCGTT TGTCAACCGC GTAATATGGG AAAACTTTTA CAAGCCCATC GTGTACATTG GCACTGACTC CGGTGAGGAA
Erannis_defolariaNPV174          TTGAAGAATT CATCGGTAGA GTGATTTGGG AGAACTTCTA CAAGCCCATC GTGTACATCG GTACCGATTC GGCTGAAGAA
Euplasia_luciparaGV248           TTGAATGCTT CTTCGAAGAC GTGCTCTGGC CCTACTTCTA CCGCCCTCTG GTCTACGTAG GCACCACCTC CGCCGAAATT
Euproctis_Chrysorrohea_NPVA10    TCGAATCGTT CATTAACCGC GTCATCTGGG AAAACTATTA CAAACCCCTC GTCTACATTG GTACCGATTC CGCCGAAGAA
Euproctis_similis_NPV768         TCGAATCGTT CATTAACCGC GTCATCTGGG AAAACTATTA CAAACCCCTC GTCTACATTG GTACCGATTC CGCCGAAGAA
Harrisina_brillians_GV           TTGAAGGTTT CTTCGAAGAC GTTTTATGGC CATACTTCCA CAGACCTTTA GTATATGTGG GCACCACTTC GGGTGAAATC
Harrisina_brillians_GVM2         TTGAAGGTTT CTTCGAAGAC GTTTTATGGC CATACTTCCA CAGACCTTTA GTATATGTGG GCACCACTTC GGGTGAAATC
HaSNPV                           TTGAAGATTT CATTACCAAC GTAATTTGGG AGAACTTTTA CAAACCAATT GTTTACGTAG GCACTGATTC TGCCGAAGAA
Heliconius_erato_NPV789          TTGAATCTTT CGTGGAGCGC GTAATTTGGG AAAACTTTTA CAAACCCATC GTGTACATTG GCACCGATTC GGGCGAAGAG
Homona_coffearia_GV744           TTGAAAGATT TTTTGACGAC ATTCTTTGGC CGTATTTCCA CCGTCCTCTA GTTTATGTGG GAACCACTTC CGCCGAAATC
Homona_coffearia_NPV745          TTGAAAGATT TTTTGACGAC ATTCTTTGGC CGTATTTCCA CCGTCCTCTA GTTTATGTGG GAACCACTTC CGCCGAAATC
Hyphantria_cunea_GV228           TCGAAAAAAT CTTTGACAAT TTCCTATGGC CCTATTTCTA TAGACCATTG GTTTGGGTAG GCACCACTTC TGGCGAAGTG
Hyphantria_cunea_NPV             TTGAACAGTT CGTTAACCGC GTAATTTGGG AGAACTTTTA CAAACCCATC GTGTACATTG GCACAGACTC TGGCGAGGAA
Hyphantria_cunea_NPVN3           TTGAACAGTT CGTTAACCGC GTAATTTGGG AGAACTTTTA CAAACCCATC GTGTACATTG GCACAGACTC TGGCGAGGAA
HzSNPV_genome                    TTGAAGATTT CATTACCAAC GTAATTTGGG AGAACTTCTA CAAACCAATT GTTTACGTAG GCACTGATTC TGCCGAAGAA
Idaea_virgulata_NPV402           TTGAATCCTT CGTTAACCGC GTAATCTGGG AAAACTTTTA CAAGCCCATC GTTTACATTG GCACAGATTC AGGTGAAGAG
```

```
Junonia_coenia_GV19                TCGAAGAGTT CTTCGATCAA TTCCTCTGGC CATATTTCCA CAGGCCGTTA GTGTGGGTCG GTACGACCTC GGCCGAAACC
Kotochalia_junodi_NPV205           TCGAGGAGTT CATTAACCGA GTTATTTGGG AAAACTTTTA TAAACCGATA GTGTACGTTG GCACGGACTC TGCCGAAGAA
Kotochalia_junodi_NPV32            TCGAGGAGTT CATTAACCGA GTTATTTGGG AAAACTTTTA TAAACCGATA GTGTACGTTG GCACGGACTC TGCCGAAGAA
Lacanobia_orecalcea_GVA6           TCGAATCTTT CTTCGAAGAC GTTCTGTGGC CATATTTCTA CCGCCCCCTC GTTTACGTCG GAACCACATC CGCCGAAATT
Lasiocampa_quercus_NPV202          TCGAAGAGTT CATCAATAGA GTGATCTGGG AGAACTTCTA CAAGCCGTTG GTGTACATTG GCACCGATTC CGCCGAAGAA
LdMNPV                             TCGAGCACTT CCTCAACTCG GTCATCTGGG ACAACTTCTA CAAGCCCATC GTGTACGTGG GCACGACCTC CGCCGAGGAG
Lonomia_obliqua_NPV                TCGAATCCTT TGTAAATCGT GTGATTTGGG AAAACTTTTA CAAGCCCGTA GTGTACATTG GCACCGATTC CGCTGAAGAA
Lymantria_monacha_NPVB4            TCGAGCACTT CTTCGATAGC GTCATTTGGG ACAACTTCTA CAAGCCCATC GTGTACGTGG GCACAGACTC GGCCGAAGAG
Lymantria_nanayi_NPV750            TTGAGCTCTT CCTCGACTCG GTCNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
Lymantria_obfuscata_NPV193         TCGAGCACTT CCTCAACNAG GTCATCTGGG ACAACTTCTA CAAGCCCATC GTGTACGTGG GCACCACCTC CTCCGAGGAG
Mahasena_corbetti_GV699            TCGAAGAGTT CATCAACAAG GTGATCTGGG ACAACTTTTA CAAGCCCATC GTGTACGTAG GTACCGACTC TGCCGAAGAG
Mahasena_corbetti_NPV595           TCGAAGAATT CATCAACAGG GTGATTTGGG AGAATTTCTA TAAACCTATC GTTTATGTAG GTACGGACTC CGCTGAAGAA
Mahasena_corbetti_NPV692           TCGAAGAATT CATCAACAGG GTGATTTGGG AGAATTTCTA TAAACCTATC GTTTATGTAG GTACGGACTC CGCTGAAGAA
Malacosoma_neustria_NPV            TTGAAGATTT CATCAACCGA GTGATCTGGG AAAACTTCTA CAAACCTCTC GTGTACATTG GTACCGATTC TGCCGAAGAA
Malacosoma_neustria_NPVk12         TTGAAGATTT CATCAACCGC GTGATCTGGG AGAACTTCTA TAAACCCCTA GTGTACATTG GTACCGACTC TGCGGAAGAG
Malacosoma_pluviale_NPVk3          TTGAAGATTT CATCAACCGC GTGATCTGGG AGAACTTCTA TAAACCCCTA GTGTACATTG GTACCGACTC TGCGGAAGAG
Malocosoma_disstria_NPV            TTGAAGATTT CATCAACCGA GTGATCTGGG AAAACTTTTA TAAACCTCTC GTGTACATCG GCACCGATTC GGCGGAAGAG
Mamestra_brassicae_NPV             TCGAAGAGTT CATCAACCGT GTCATATGGG AGAACTTCTA CAAGCCAATC GTGTACGTAG GCACAGATTC GGCTGAGGAA
Mamestra_brassicae_NPV274          TCGAACAGTT CATCGAYCGT GTCATCTGGG AGAACTTCTA CAAGCCCATC GTGTACGTCG GTACCGACTC TGCTGAAGAG
Mamestra_brassicae_NPVK9           TCgAAGAGTT CATCAACCGT GTCATATGGG AGAACTTCTA CAAGCCAATC GTGTACATAG GCACAGATTC GGCTGAGGAA
Mamestra_configurata_NPVA          TCGAAGAGTT CATCAACCGT GTCATCTGGG AGAACTTCTA CAAGCCCATC GTGTACGTTG GTACCGATTC GGCTGAGGAA
Mamestra_configurata_NPVB          TCGAAGAGTT CATCAACCGT GTCATATGGG AGAACTTCTA CAAGCCAATC GTGTACGTAG GCACAGATTC GGCTGAGGAA
Melanchra_persicariaGV26           TCGAATCTTT CTTCGAAGAC GTTCTGTGGC CATACTTCCA CCGCCCCCTA GTTTACGTTG GAACCACCTC CGCCGAAATC
Mynthimna_pseudolatia_NPVA1        TTGAAGAATT CATCAACCGC GTCATCTGGG AGAACTTCTA CAAGCCGATC GTGTACGTCG GTACCGATTC GGCCGAAGAA
'Mynthimna separata (Lese)NPV'     TCGAAGAGTT CATCAACCGT GTCATCTGGG AGAACTTCTA CAAGCCAATC GTGTACGTAG GCACAGATTC GGCTGAGGAA
Mynthimna_separata_ChinaNPVA3      TCGAGAGCTT CATCGACAAA GTCATCTGGT ACAACTTCTA CAAACCGATC GTGTACGTTG GAACCGATTC GGCCGAAGAG
Mynthimna_separata_IndiaNPVA4      TCGAGAGCTT CATCGACAAA GTCATCTGGT ACAACTTCTA CAAACCGATC GTGTACGTTG GAACCGATTC GGCCGAAGAG
Mynthimna_separata_Japan_NPVA2     TCGAAGAGTT CATCAACCGT GTCATCTGGG AGAACTTCTA CAAGCCAATC GTGTACGTAG GCACAGATTC GGCTGAGGAA
Natada_nararia_GV254               TTGAAAGGTT CTTTGATGAT GTGCTATGGC CATATTTCCA CAGACCTTTA GTTTATGTAG GAACCACTTC GGCAGAAATA
Nymphalis_io_NPV353                TCGAAGAGTT CATTTCTAAA GTAATTTGGG AGAACTTCTA CAAACCAATC GTGTACCTCG GTACCGATTC CGCTGAAGAA
Operoptera_brumata_NPV291          TCGAAGAGTT CATCGAGAAG GTGATCTGGG ACAACTTTTA CAAGCCCCTC GTGTACGTCG GTACCGACTC TGCCGAAGAG
Operoptera_brumata_NPVi4           TCGAAGAGTT CATCGAGAAG GTGATCTGGG ACAACTTTTA CAAGCCCCTC GTGTACGTCG GTACCGACTC TGCCGAAGAG
Operoptera_brumata_NPVi5           TCGAAGAGTT CATCGAGAAG GTGATCTGGG ACAACTTTTA CAAGCCCCTC GTGTACGTCG GTACCGACTC TGCCGAAGAG
OpMNPV                             TTGAATCGTT CGTAAACCGC GTCATCTGGG AGAACTTTTA TAAGCCCATC GTGTACATTG GCACGGATTC GAGCGAGGAG
OpSNPV                             TTGAGGAGTT TATTAACCGC GTTCATTGGG AGAACTTCTA CAAACCCATC GTGTACGTCG GCACAGATTC TGCCGAGGAA
Orgya_antiqua_NPVA9                TTGAGGAGTT TATTAACCGC GTCATTTGGG AGAACTTCTA CAAACCCATC GTGTACGTCG GCACAGATTC TGCCGAGGAA
Orgya_leucostigma_NPVB6            TTGAAGAATT TATTAACCGC GTCATTTGGG AAAACTTTTA CAAACCCATT GTGTATGTAG GAACAGATTC CGCCGAAGAA
Orgya_mixta_NPV67                  TTGAAGAGTT TATCAACCGC GTTATCTGGG AGAACTTCTA CAAGCCCATC GTGTACGTGG GAACCGATTC GGCCGAGGAA
Orgya_pseudotsugattaNPVB5          TTGAAGAATT CATTAACCGC GTCATTTGGG AAAACTTTTA CAAACCCATT GTGTACATAG GAACAGATTC CACCGAAGAG
Orgyia_anartoides_NPV              TCGAAGAATT TATCAACCGC GTGATTTGGN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
Pachypasa_papyri_NPV436            TTGAAGATTT CATCAACAAA GTCATTTGGG AGAACTTCTA CAAACCCATA GTTTACATTG GTACCGATTC TGCCGAAGAA
Panolis_flammea_MNPV               TTGAAGAGTT CATCAACCGT GTCATCTGGG AGAACTTCTA CAAGCCCATC GTGTACGTTG GTACCGATTC GGCTGAGGAA
Panolis_flammea_NPVi3              TCGAAGAGTT CATCAACAGT GTCATCTGGG AGAACTTCTA CAAGCCCATC GTGTACGTTG GTACCGATTC GGCTGAGGAA
Perina_nuda_NPV                    TTGAATCGTT CGTAAACCGC GTCATCTGGG AAAACTTTTA CAAGCCCATT GTGTACATTG GCACGGATTC GGGCGAGGAG
PhopGV                             TTGAAACGTT CTTTGACGAC GTCCTGTGGC CATACTTCTA CCGTCCTCTG GTCTACGTGG GCACCACATC GTCCGAAATC
Phthorimaea_operculella_GVC3       TTGAAACGTT CTTTGACGAC GTCCTGTGGC CATACTTCTA CCGTCCTCTG GTCTACGTGG GCACCACATC GTCCGAAATC
Phthorimaea_operculella_GVC4       TTGAAACGTT CTTTGACGAC GTCCTGTGGC CATACTTCTA CCGTCCTCTG GTCTACGTGG GCACCACATC GTCCGAAATC
Pieris_brassicae_GV                TTGAAACTTT CTTTGAAGAT GTACTATGGC CATACTTCCA CAGACCACTC GTCTACATCG GTACCACCTC TAGCGAAACC
Pieris_brassicae_GV384             TTGAAACTTT CTTTGAAGAT GTACTATGGC CATACTTCCA CAGACCACTC GTCTACATCG GTACCACCTC TAGCGAAACC
Pieris_rapaeGV95                   TTGAAACTTT CTTTGAAGAT GTACTATGGC CATACTTCCA CAGACCACTC GTCTACATCG GTACCACCTC TAGCGAAACC
Plodia_interpunctella_GVB3         TCGAAACCTT CTTCGACGAC GTCTTATGGC CATACTTCCA CAGGCCCCTA GTGTACGTGG GCACTACCTC CTCCGAGATG
Plusia_gammaNPV287                 TCGAACAGTT CATCGATCGT GTCATCTGGG AGAACTTCTA CAAGCCCATC GTTTACATCG GTACCGACTC TGCTGAAGAG
Plusia_orichalcea_NPV              TTGAGGAGTT CATCAACCGT GTGATCTGGG AGAACTTCTA CAAACCCATT GTTTATGTAG GCACCGACTC TGCCGAAGAA
'Polygonia c-album NPV380'         TCGAAGAGTT CATTTCTAAA GTAATKTGGG AGAACTTCTA CAAACCAATC GTGTACCTCG GTACCGATTC CGCTGAAGAA
PxGV                               TCGAGACGTT CTTCGAGGAC AACCTGTGGC CTTACTTTCA CCGTCCTCTG GTGTACGTGG GTACCACTTC CGGAGAGGTT
RaouNPV                            TCGAGTCGTT TGTGAGCCGC GTCATATGGG AGAATTTCTA CAAACCCATC GTTTACATTG GCACAGACTC TGGCGAAGAA
Samia_ricini_NPV28                 TCGAACAGTT CATCGATCGT GTCATCTGGG AGAACTTCTA CAAGCCCATC GTTTACATCG GTACCGACTC TGCTGAAGAG
Shpinx_ligustri_NPV334             TCGAGTCGTT TGTGAACCGC GTCATATGGG AGAACTTCTA CAAACCCATC GTTTACATCG GCACAGACTC TGGCGAAGAA
Smerinthus_ocellata_NPV819         TTGAGAGTTT CATCGACAAG GTGATATGGT ACAACTTTTA CAAGCCCATC GTGTACGTGG GCACCGATTC GGCCGAAGAG
Sphinx_ligustri_NPV362             TCGAACAGTT CATCGATCGT GTCATCTGGG AGAACTTCTA CAAGCCCATC GTTTACATCG GTACCGACTC TGCTGAAGAG
Spilosoma_lutea_GV126              TCGAAAAATT CTTCGACAAC TTCCTGTGGC CATACTTCTA CCGTCCCCTC GTATGGGTGG GCACCACTTC TGGTGAAGTA
Splisoma_imparilis_NPVN2           TTGAACAGTT CGTTAACCGC GTAATTTGGG AGAACTTTTA CAAACCCATC GTGTACATTG GCACAGACTC TGGCGAGGAA
Spodoptera__exempta_NPVK11         TCGAAGAGTT CATCAACCGT GTCATCTGGG AGAACTTCTA CAAACCCATC GTGTACATCG GTACCGATTC CGGAGAAGAG
Spodoptera_exigua_MNPV             TCGAGGAATT CATCAACCGC GTCATTTGGG AGAACTTTTA CAAACCCATC GTTTACGTAG GAACCGACTC GGGTGAGGAA
Spodoptera_frugiperda_NPV          TCGAAGAGTT CATCAACCGT GTCATCTGGG AGAACTTCTA CAAACCCATC GTCTACGTAG GAACCGACTC TGGTGAGGAA
Spodoptera_frugiperda_NPVK10       TCGAAGAGTT CATCAACCGT GTCATCTGGG AGAACTTCTA CAAACCCATC GTCTACGTAG GAACCGACTC TGGTGAGGAA
Spodoptera_litoralis_NPV           TTGAGAGTTT CATCGACAAG GTGATATGGT ACAACTTTTA CAAGCCCATC GTGTACGTGG GCACCGATTC GGCCGAAGAG
Spodoptera_littoralis_GV66         TTGAAGAGTT CTTTGACAGC GTGCTGTGGC CATACTTCCA TCGTCCCTTG GTCTATGTGG GAACCACTTC CGCCGAGATC
Spodoptera_littoralis_NPV440       TTGAGAGTTT CATCGACAAG GTGATATGGT ACAACTTTTA CAAGCCCATC GTGTACGTGG GCACCGATTC GGCCGAAGAG
Spodoptera_litura_GV766            TTGAAGAGTT CTTTGACAGC GTGCTGTGGC CATACTTCCA TCGTCCCTTG GTCTATGTGG GAACCACTTC CGCCGAGATC
Spodoptera_litura_NPVg             TCGAAGAATT CATCAACCGC GTCATCTGGG AAAACTTCTA CAAACCCATC GTGTACGTAG GAACCGACTC CGGTGAAGAG
Spodoptera_litura_NPV2             TCGAGAGCTT CATCGACAAG GTGATATGGT ACAACTTTTA CAAGCCCATC GTGTACGTGG GCACCGACTC GGCTGAAGAG
Spodoptera_sunia_NPVK8             TTGAAGAGTT TATTGCTCGC GTGATCTGGG AGAACTTCTA CAAGCCCATC GTTTACGTAG GAACCGATTC CGCCGAGGAA
Spodoptera_sunia_NPVK7             TTGAGGAATT TATCAACCGT GTAATTTGGG AAAACTTTTA TAAACCCaTC GTGTACGTTG GAACCGATTC GGGCGAGGAA
Thysanoplusia_orichalcea_NPV       TTGAGTCGTT TGTGAACCGT GTGATTTGGG AAAATTTCTA CAAACCTATC GTTTACATTG GTACTGATTC TGGCGAAGAG
Thysanoplusia orichalcea NPV2      TTGAGgagTT caTcAACCGT GTGATcTGGG AgAAcTTCTA CAAACCcATt GTTTAtgTaG GcACcGAcTC TGccGAAGAa
Trabala_vishnou_NPV195             TTGAAGAATT CATCAACCGT GTCATTTGGG AAAATTTCTA TAAACCTATT GTGTACGTTG GCACCGATTC CGCCGAAGAA
Trichoplusia_ni_GV                 TTGAAGAGTT CTTCGATCAG GTTCTGTGGC CATACTTCCA CCGTCCTCTC GTCTACGTGG GCACGACGTC TGCCGAAATT
Trichoplusia_ni_MNPVC2             TCGAACAGTT CATCGATCGT GTCATCTGGG AGAACTTCTA CAAGCCCATC GTTTACATCG GTACCGACTC TGCTGAAGAG
Trichoplusia_ni_SNPV               TTGAAGAGTT TATTGCTCGC GTGATCTGGG AAAACTTCTA CAAACCCATA GTTTACGTAG GAACCGATTC CGCCGAGGAA
Trichoplusia_orichalceaNPVB9       TCGAGGAGTT CATCAGCCGT GTGATCTGGG AGAACTTCTA CAAACCCATC GTTTACGTAG GAACCGACTC TGCCGAAGAG
Utetheisa_pulchella_NPV142         TCgAAgAGTT CATCGACAAG GTGATCTGGG ACAACTTTTA CAAGCCCCTC GTGTACATTG GTACCGACTC TGGCGAAGAG
Vanessa_alatanta_NPV54             TCGAAGAGTT CATTTCTAAA GTAATTTGGG AGAACTTCTA CAAACCAATC GTGTACCTCG GTACCGATTC CGCTGAAGAA
Vanessa_cardui_NPV367              TCGAAGAGTT CATTTCTAAA GTAATTTGGG AGAACTTCTA CAAACCAATC GTGTACCTCG GTACCGATTC CGCTGAAGAA
Wiseana_cervinata_GV342            TTGAAAAGTT CTTTGAGGAC GGTCTGTGGC CATATTTCCA TCGTCCTTTG GTGTATGTAG GAACGACGTC CGGAGAGACG
Wiseana_cervinata_NPV344           TCGAGCACTT CATCAACAAG GTAATTTGGG AGAATTTCTA CAAGCCAATC GTTTACATAG GAACCGACTC TGCGGAGGAC
Wiseana_signata_NPV                TCGAGCACTT CATCAACAAG GTAATTTGGG AGAATTTCTA CAAACCAATA GTTTACATAG GAACCGACTC TGCGGAGGAC
XcGV                               TTGAGGAGTT CTTCGATCAG GTCCTGTGGC CATACTTCCA CCGTCCCCTG GTGTACGTGG GCACAACGTC TGCCGAGATT

Abraxasgrossulariata_NPV112        GAGGAAATCC TCATTGAGGT GTCGCTCGTG TTCAAGGTGA AGGAGTTCGC CCCCGAC
Achaea_faber_NPV64                 GAGGAAATTC TCTTGGAACT GTCTCTGGTG TTCAAGATTA AGGAGTTCGC CCCCGAC
AchaeajanataGV835                  GAAGAAGTCA TGATTGAAGT GTCTCTTCTG TTCAAGATTA AGGAGTTCGC CCNNNNN
AcMNPV                             GAGGAAATTC TCCTTGAAGT TTCCCTGGTG TTCAAAGTAA AGGAGTTTGC ACCAGAC
AcMNPVK1                           GAGGAAATTC TCCTTGAAGT TTCCCTGGTG TTCAAAGTAA AGGAGTTCGC CCCCGAC
Actias_selene_NPV47                GAGGAAATTC TCATCGAGGT TTCGCTTGTG TTCAAGGTCA AGGAGTTCGC CCCCGAC
Adoxophyes_orana_GV                GAAGAGATTA TGATTGAGGT GTCGCTTCTG TTTAAGATCA AAGAATTCGC ACCCGAC
AfMNPVK2                           GAGGAAATCC TAATTGAGGT TTCTCTCGTT TTCAAAGTAA AGGAGGGCGC CCCCGAC
Aglais_urticae_NPVA7               GAGGAAATCT TACTCGAAGT ATCATTGGTG TTCAAGATCA AGGAGTTCGC CCCGACA
Agraulis_vanillare_NPV779          GAGGAAATTC TGATTGAGGT GTCGCTCGTG TTCAAAGTAA AGGAGTTCGC CCCCGAc
Agrotis_segetum_NPVi6              GAAGAAATCC TGCTCGAGCT GTCGCTCATT TTCAAGATCA AGGAGTTCGC CCNNNNN
Agrotis_segetum_NPVK4              GAAGAAATAC TGCTCGAGCT GTGGCTCATT TTCAAGATCA AGGAGGGCGC CCCCGAC
Ammalo_insulata_NPV175             GAAGAAATTT TAATCGAGGT TTCGCTTGTG TTTAAGGTCA AGGAGTTCGC CCCCGAC
Amonis_sabulifera_NPV14            GAGGAAATCC TTCTCGAAGT AGCGCTCGTG TTCAAAATTA AGGAGTTCGC CCCCGAC
Amsacta_albistriga_NPV             GAGGAGATAC TTATTGAAGT GTCACTCGTG TTTAAAGTGA AAGAGTTTGC GCCCGAC
```

```
Antherea_pernyi_NPV                GAGGAAATTC TCATCGAGGT TTCGCTTGTG TTCAAGGTCA AGGAGTTTGC GCCCGAC
Antherea_polyphemus_NPV30          GAGGAAATCC TAATTGAGGT TTCTCTCGTT TTCAAAGTAA AGGAGTTCGC CCNNNNN
Anthophila_fabricana_GV790         GAAGAAATCA TGATCGAAGT GTCTCTTCTG TTCAAAATCA AGGAGTTCGC CCCCGAC
Anticarsia_gemnitalis_NPV          GAAGAAATTC TTATTGAAGT CTCGCTCGTG TTCAAGGTGA AAGAGTTTGC GCCCGAC
Aporia_crataegi_NPV106             GAAGAGGTTT TGCTTGAAGT GTCTTTAATT TTCAAAGTTA AGGAGTTCGC CCCCGAC
Archips_cerasivora_NPV             GAGGAAATCC TCATCGAGGT CTCGCTGGTG TTTAAGGTGA AAGAGTTTGC GCCCGAC
Aroa_discalis_NPV63                GAAGAAATCC TGCTGGAAGT ATCGCTCGTC TTCAAAATTA AGGAGTTCGC CCCCGAC
Artogeia_napi_GV242                GAAGAAATCT TACTCGAAGT ATCTTTCTTG TTCAAGATCA AGGAGTTCGC CCCNNNN
Attacus_ricini_NPV                 GAGGAAATTC TCATCGAGGT TTCGCTTGTG TTCAAGGTCA AGGAGTTCGT GCCCGAC
BmNPV                              GAGGAAATCC TAATTGAGGT TTCTCTCGTT TTCAAAATAA AGGAGTTTGC ACCAGAC
Bombyx_mori_NPV460                 GAGGAAATCC TAATTGAGGT TTCTCTCGTT TTCAAAATAA AGGAGTTCGC CCCGACA
Buzura_edwardsii_NPV505            GAGGAGATCC TCCTTGAACT GTCGCTCGTG TTAAAGATCA AGGAGTCCGC CCCCGAC
Buzura_Supressaria_NPV             GAAGAAATTC TTCTTGAAGT TTCTTTGGTA TTTAAAGTTA AAGAATTTGC GCCTGAT
Cacoecia_murinana_GV37             GAAGAAATTA TGATCGAGGT GTCGCTATTG TTTAAAATTA AGGAGTTCGC CCCGACA
Catopsila_pomona_NPV78             GAAGAAATTT TAATCGAAGT GTCGCTCGTG TTTAAAGTCA AGGAGTTCGC CCCGACA
Cerapteryx_gramini_NPVA5           GAGGAGATCC TGCTCGAACT TTCTCTCGTG TTCAAGATCA AGGAGTTCGC CCCCGAC
ChfuNPV                            GAGGAAATCC TCATTGAGGT TTCGCTGGCT NNNNNNNNNN NNNNNNNNNN NNNNNNN
ChfuNPVDEF                         GAAGAGATTC TTATTGAAGT CTCGCTCGTG TTCAAAGTGA AGGAGTTTGC GCCTGAC
Choristoneura_furamifera_GV        GAAGAAATTA TGATCGAGGT GTCGCTATTG TTTAAAATTA AAGAATTCGC ACCAGAT
Choristoneura_rosaceana_NPV        GAGGAAATTC TCATTGAGGT TTCGCTGGTG TTCAAGGTTA AAGAGTTTGC GCCTGAT
Cirphis_unipuncta_NPV172           GAAGAAATCT TACTAGAGGT TGCTCTGGTG TTCAAAATCA AGGAGGTCGC CCCCGAC
CpGV                               GAAGAGATCA TGATTGAGGT GTCGCTCTTG TTCAAGATCA AGGAGTTTGC ACCCGAC
Cryptophlebia_leucotreta_GV        GAGGAGATCA TGATTGAAGT TTCTCTGTTG TTCAAGATCA AGGAGTTTGC ACCCGAC
Cynosarga_chrysolopa_NPV302        GAAGAAATAC TTATCGAAGT TTCTCTCGTG TTTAAAGTTA AGGAGTGCGC CCCCGAC
Darna_trima_GV545                  GAAGAAATTA TGATTGAAGT TTCGCTGTTG TTCAAAATTA AGGAGTTCGC CCCCGAC
Diacrisia_obliqua_NPV798           GAAGAAATTT TAATTGAAGT CTCTCTCGTG TTCAAAGTAA AGGAGTTCGC CCCCGAC
Diatrea_saccharalis_NPVK6          GAGGAAATCA TGATTGAGGT ATCTCTAATC TTCAAGATTA AGGAGTTCGC CCCCGAC
Ectropis_obliqua_NPV               GAGGAAATCC TACTTGAAGT TTCGCTTGTC TTCAAAATCA AAGAATTCGC CCCTGAC
Ephestia_cautella_NPVC5            GAGGAAATTC TTATTGAAGT TTCGCTTGTA TTCAAAGTGA AGGAGTTCGC CCCCGAC
EppoMNPV                           GAGGAAATCC TGATCGAGGT GTCGCTCGTG TTCAAAGTAA AAGAGTTTGC GCCTGAC
Erannis_defolariaNPV174            GAGGAAATTC TGTTAGAGGT TTCTCTTGTG TTCAAAATTA AGGAGTTCGC CCCCGAC
Euplasia_luciparaGV248             GAAGAAATCA TGATCGAAGT GTCTCTGCTG TTCAAGATCA AGGAGTTCGC CCCCGAC
Euproctis_Chrysorrohea_NPVA10      GAAGAGATTC TCATTGAAGT GTCTTTGGTG TTTAAAGTAA AGGAGTTCGC CCCCGAC
Euproctis_similis_NPV768           GAAGAGATTC TCATTGAAGT GTCTTTGGTG TTCAAAGTAA AGGAGTTCGC CCCGACA
Harrisina_brillians_GV             GAAGAAATTA TGATCGAAGT GTCTTTACTA TTTAAAATTA AAGAGTTTGC ACCTGAT
Harrisina_brillians_GVM2           GAAGAAATTA TGATCGAAGT GTCTTTACTA TTTAAAATTA AGGAGTTCGC CCCCGAC
HaSNPV                             GAGGAAATAC TCCTAGAAGT TTCTTTGATA TTTAAGATCA AAGAATTTGC ACCTGAC
Heliconius_erato_NPV789            GAGGAAATTC TGATTGAGGT CTCGCTCGTG TTCAAAGTAA AGGAGTTCGC CCCCGAC
Homona_coffearia_GV744             GAAGAAATTA TGATTGAGGT GTCGCTGTTG TTTAAAATTA AGGAGTTCGC CCCCGAC
Homona_coffearia_NPV745            GAAGAAATTA TGATTGAGGT GTCGCTGTTG TTTAAAATTA AGGAGTTCGC NNNNNNN
Hyphantria_cunea_GV228             GAGGAAATTT TGATTGAAGT TTCTCTTTTA TTCAAAATTA AGGAGTTCGC CCCCGAC
Hyphantria_cunea_NPV               GAAGAAATTT TAATTGAAGT CTCTCTCGTG TTCAAAGTAA AGGAGTTTGC ACCCGAC
Hyphantria_cunea_NPVN3             GAAGAAATTT TAATTGAAGT CTCTCTCGTG TTCAAAGTAA AGGAGTTCGC CCCCGAC
HzSNPV_genome                      GAGGAAATAC TCCTAGAGGT TTCTTTGATA TTTAAGATCA AAGAATTTGC ACCTGAC
Idaea_virgulata_NPV402             GAAGAAATCC TCATCGAGGT CTCGCTCGTG TTTAAGGTAA AGGAGTTCGC CCNNNNN
Junonia_coenia_GV19                GAGGAAATTA TGATCGAAGT GTCGTTACTG TTCAAGATTA AGGAGTTCGC CCCCGAC
Kotochalia_junodi_NPV205           GAGGAGATTC TGATGGAAGT GTCGTTGGTG TTCAAAATTA AGGAGTTCGC CCCGCAT
Kotochalia_junodi_NPV32            GAGGAGATTC TGATGGAAGT GTCGTTGGTG TTCAAAATTA AGGAGTTCGC CCCGACA
Lacanobia_orecalcea_GVA6           GAAGAGATCA TGATCGAAGT TTCGCTTCTC TTCAAGATCA AGGAGTTCGC CCCGACA
Lasiocampa_quercus_NPV202          GAGGAAATTC TGTTGGAAGT GTCTCTAGTG TTTAAAATCA AGGAGTTCGC CCCCGAC
LdMNPV                             GAGGAGATCC TGCTCGAGGT CTCCTTGGTG TTCAAGATCA AGGAGTTCGC TCCCGAC
Lonomia_obliqua_NPV                GAAGAAATTC TCATCGAAGT GTCNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNN
Lymantria_monacha_NPVB4            GAGGAGATCC TGCTCGAGGT GTCGTTGGTG TTCAAGATCA AGGAGTTCGC CCCCGAC
Lymantria_nanayi_NPV750            NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNN
Lymantria_obfuscata_NPV193         GAGGAGATCC TGCTCGAGGT CTCCTTGGTG TTCAAGATCA AGGAGGTCGC CCCCGAC
Mahasena_corbetti_GV699            GAGGAGATTC TTATTGAAGT GTCACTGGTG TTCAAAATCA AGGAGTTCGC CCCCGAC
Mahasena_corbetti_NPV595           GAGGAGATTC TGATGGAGGT GTCATTAGTG TTTAAAATTA AGGAGTTCGC CCCCGAC
Mahasena_corbetti_NPV692           GAGGAGATTC TGATGGAGGT GTCATTAGTG TTTAAAATTA AGGAGGGCGC CCCCGAC
Malacosoma_neustria_NPV            GAGGAAATTC TCTTAGAGGT GTCTTTGGTG TTCAAAATCA AAGAGTTCGC TCCCGAC
Malacosoma_neustria_NPVk12         GAGGAAATTC TTTTAGAAGT GTCTTTGGTG TTTAAAATCA AGGAGTTCGC CCCCGAC
Malacosoma_pluviale_NPVk3          GAGGAAATTC TTTTAGAAGT GTCTTTGGTG TTTAAAATCA AGGAGTTCGC CCCCGAC
Malocosoma_disstria_NPV            GAAGAAATTC TTTTAGAAGT ATCTTTAGTG TTTAAAATCA AAGAATTCGC TCCCGAC
Mamestra_brassicae_NPV             GAGGAAATTC TTCTCGAGGT TTCTCTTGTG TTCAAAATCA AAGAGTTTGC GCCTGAT
Mamestra_brassicae_NPV274          GAGGAAATTC TCCTTGAAGT TTCCCTGGTG TTCAAAGTCA AGGAGTTCGC CCCCGAC
Mamestra_brassicae_NPVK9           GAGGAAATTC TTCTCgAGGT TTCTCTTGTG TTCAAAATCA AGGAGTTCGC CCCCGAC
Mamestra_configurata_NPVA          GAAGAAATCC TCCTAGAGGT TGCTCTGGTG TTCAAAATCA AGGAGTTTGC GCCTGAT
Mamestra_configurata_NPVB          GAGGAAATTC TTCTCGAGGT TTCTCTTGTG TTCAAAATCA AAGAGTTTGC GCCTGAT
Melanchra_persicariaGV26           CAAGAGATCA TGATCGAAGT T-GGCTCCTC TTCAAGATCA AGGAGTTCGC CCCCGAC
Mynthimna_pseudolatia_NPVA1        GAGGAGATCC TACTCGAGGT CTCGCTCGTC TTCAAGATCA AGGAGTTCGC CCCCGAC
'Mynthimna separata (Lese)NPV'     GAAGAAATTC TTCTCGAGGT TTCTCTTGTG TTCAAAATCA AAGAGTTTGC GCCTGAT
Mynthimna_separata_ChinaNPVA3      GAGGAAATCC TGCTTGAAGT CTCTCTCGTG TTCAAGATCA AGGAGTTCGC CCCCGAC
Mynthimna_separata_IndiaNPVA4      GAGGAAATCC TGCTTGAAGT CTCTCTCGTG TTCAAGATCA AGGAGTTCGC CCCCGAC
Mynthimna_separata_Japan_NPVA2     GAGGAAATTC TTCTCGAGGT TTCTCTTGTG TTCAAAATCA AGGAGTTCGC CCCCGAC
Natada_nararia_GV254               GAAGAAATTA TGATTGAAGT TTCTCTTTTG TTCAAAATTA AGGAGTTCGN NNNNNNN
Nymphalis_io_NPV353                GAGGAAATCT TACTCGAAGT ATCATTGGTG TTCAAGATCA AGGAGTTCGC CCCCGAC
Operoptera_brumata_NPV291          GAGGAGATTC TTCTTGAAGT CAGTCTGGTG TTCAAAATCA AGGAGTTCGC CCCSACA
Operoptera_brumata_NPVi4           GAGGAGATTC TTCTTGAAGT CAGTCTGGTG TTCAAAATCA AGGTGTTCGC CCCGACA
Operoptera_brumata_NPVi5           GAGGAGATTC TTCTTGAAGT CAGTCTGGTG TTCAAAATCA AGGAGTTCGC CCCGACA
OpMNPV                             GAGGAAATTC TCATCGAGGT GTCGCTCGTG TTTAAGGTAA AAGAGTTTGC GCCCGAC
OpSNPV                             GAAGAAATAT TACTTGAAGT TTCGCTTGTG TTCAAAATTA AGGAATTTGC ACCAGAC
Orgya_antiqua_NPVA9                GAAGAAATAT TACTTGAAGT TTCGCTTGTG TTCAAAATTA AGGAGTTCGC CCCCGAC
Orgya_leucostigma_NPVB6            GAAGAAATCT TACTTGAGGT TTCTCTTGTG TTTAAAATTA AGGAGTTCGC CCCCGAC
Orgya_mixta_NPV67                  GAAGAAATCC TGCTGGAAGT ATCGCTCGTC TTCAAAATTA AGGAGTTCGC CCCCGAC
Orgya_pseudotsugattaNPVB5          GAGGAAATCc TCATCGAGGT GTCTCTCGTG TTTAAAATAA AGGAGTTCGC CCCCGAC
Orgyia_anartoides_NPV              NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNN
Pachypasa_papyri_NPV436            GAAGAAATTC TTTTAGAGGT TTCGCTGGTG TTTAAAATAA AGGAGTTCGC CCCCGAC
Panolis_flammea_MNPV               GAAGAAATCC TCCTAGAGGT TTCTCTGGTG TTTAAAATCA AGGAGTTTGC GCCTGAT
Panolis_flammea_NPVi3              GAAGAAATCC TCCTAGAGGT TTCTCTGGTG TTTAAAATCA AGGAGTTCGC CCCCGAC
Perina_nuda_NPV                    GAAGAAATTC TCATCGAAGT GTCGTTGGTG TTTAAGGTAA AAGAGTTTGC GCCCGAC
PhopGV                             GAAGAAATCA TGATCGAGGT TTCGCTGTTG TTCAAAATCA AGGAATTCGC GCCCGAC
Phthorimaea_operculella_GVC3       GAAGAAATCA TGATCGAGGT TTCGCTGTTG TTCAAAATCA AGGAGTTCGC CCCCGAC
Phthorimaea_operculella_GVC4       GAAGAAATCA TGATCGAGGT TTCGCTGTTG TTCAAAATCA AGGAGTTCGC CCCGACA
Pieris_brassicae_GV                GAAGAAATCT TACTCGAAGT ATCTTTCTTG TTCAAGATCA AGGAATTCGC ACCCGAT
Pieris_brassicae_GV384             GAAGAAATCT TACTCGAAGT ATCTTTCTTG TTCAAGATCA AGGAGTTCGC CCCCGAC
Pieris_rapaeGV95                   GAAGAAATCT TACTCGAAGT ATCTTTCTTG TTCAAGATCA AGGAGTTCGC CCCCGAC
Plodia_interpunctella_GVB3         GAGGAGATCA TGATCGAGGT ATCTCTGCTG TTCAAGATCA AGGAGTTCGC CCCCGAC
Plusia_gammaNPV287                 GAGGAAATTC TCCTTGAAGT TTCCCTGGTG TTCAAAGTAA AGGAGTTCGC CCCGCNN
Plusia_orichalcea_NPV              GAGGAAATCC TTCTTGAGGT GTCTTTAGTG TTCAAAATCA AAGAGTTCGC CCCAGAC
'Polygonia c-album NPV380'         GAGGAAATCT TACTCGAAGT ATCATTGGTG TTCAAGATCA AGGAGTTCGC CCCCGAC
PxGV                               GAGGAGATTA TGATTGAGGT CGCGCTTATC TTCAAAATCA AGGAATTCGC GCCCGAC
RaouNPV                            GAGGAAATCC TAATTGAGGT TTCTCTCGTT TTCAAAGTAA AGGAGTTTGC ACCAGAC
Samia_ricini_NPV28                 GAGGAAATTC TCCTTGAAGT TTCCCTGGTG TTCAAAGTAA AGGAGTTCGC CCCCGAC
Shpinx_ligustri_NPV334             GAGGAAATCC TAATTGAGGT TTCTCTCGTT TTCAAAGTAA AGGAGTTCGC CCCCGAC
Smerinthus_ocellata_NPV819         GAGGAGATCC TTCTCGAAGT GTCGCTCGTG TTCAAGATCA AGGAGTTCGC CCCCGAC
Sphinx_ligustri_NPV362             GAGGAAATTC TCCTTGAAGT TTCCCTGGTG TTCAAAGTAA AGGAGTTCGC CCCCGAC
```

```
Spilosoma_lutea_GV126              GAAGAAATCT TAATTGAAGT GTCGCTTCTG TTTAAGGTGA AGGAGTTCGC CCCCGAC
Splisoma_imparilis_NPVN2           GAAGAAATTT TAATTGAAGT CTCTCTCGTG TTCAAAGTAA AGGAGTTCGC CCCCGAC
Spodoptera__exempta_NPVK11         GAGGAAATTC TCTTGGAACT GTCTCTGGTG TTCAAGATTA AGGAGTTCGC CCCCGAC
Spodoptera_exigua_MNPV             GAGGAAATCC TCCTCGAACT GTCGCTCGTG TTCAAGATCA AGGAATTCGC ACCCGAT
Spodoptera_frugiperda_NPV          GAGGAGATCC TCCTTGAACT GTCGCTCGTG TTCAAGATCA AGGAGTTTGC TCCTGAC
Spodoptera_frugiperda_NPVK10       GAGGAGATCC TCCTTGAACT GTCGCTCGTG TTCAAGATCA AGGAGTTCGC CCCCGAC
Spodoptera_litoralis_NPV           GAGGAGATCC TTCTCGAAGT GTCGCTCGTG TTCAAGATCA AAGAGTTTGC TCCCGAC
Spodoptera_littoralis_GV66         GAAGAGGTCA TGATCGAAGT AGCGCTGCTG TTTAAGATCA AGGAGTTCGC CCCCGAC
Spodoptera_littoralis_NPV440       GAGGAGATCC TTCTCgAAGT GTCGCTCGTG TTCAAGATCA AGGAGTTCGC CCCCGAC
Spodoptera_litura_GV766            GAAGAGGTCA TGATCGAAGT AGCGCTGCTG TTTAAGATCA AGGAGTTCGC CCCCGAC
Spodoptera_litura_NPVg             GAAGAAATCC TCCTCGAAGT GGCCCTCGTT TTCAAGATCA AAGAGTTTGC GCCCGAC
Spodoptera_litura_NPV2             GAGGAGATCC TTCTCGAAGT GTCGCTCGTG TTCAAGATCA AAGAGTTTGC TCCCGAC
Spodoptera_sunia_NPVK8             GAGGAGATTC TTCTTGAAGT GTCTTTAGTC TTCAAGATTA AGGAGTTCGC CCCCGAC
Spodoptera_sunia_NPVK7             GAGGAGATAC TCCTCgAACT GGCTGTCGTG TTCAAGATCA AGGAGTTCGC CCCCGAC
Thysanoplusia_orichalcea_NPV       GAGGAGATTC TTATTGAAGT TTCTCTTGTT TTCAAAGTAA AGGAGTTTGC GCCAGAC
Thysanoplusia orichalcea NPV2      GAGGAaATcC TTcTTGAgGT gTCTtTaGTg TTCAAAaTcA AaGAGTTcGC cCCAGAC
Trabala_vishnou_NPV195             GAGGAGATTC TACTTGAGGT CTCGCTCGTG TTCAAGATCA AGGAGTTCGC CCCCGAC
Trichoplusia_ni_GV                 GAAGAGGTCA TGATCGAAGT GGCTCTGCTC TTCAAGATCA AAGAATTCGC TCCTGAT
Trichoplusia_ni_MNPVC2             GAGGAAATTC TCCTTGAAGT TTCCCTGGTG TTCAAAGTAA AGGAGTTCGC CCCCGAC
Trichoplusia_ni_SNPV               GAGGAGATTC TTCTTGAAGT GTCTTTAGTC TTTAAAATTA AGGAATTCGC TCCCGAC
Trichoplusia_orichalceaNPVB9       GAGGAGATTC TCCTCGAGGT GTCTCTCGTG TTCAAAATCA AGGAGTTCGC CCCCGAC
Utetheisa_pulchella_NPV142         GAAGAGATTC TTCTCGAAGT CACGCTGGTG TTTAAAGTCA AGGAGTTCGC CCCCGAC
Vanessa_alatanta_NPV54             GAGGAAATCT TACTCGAAGT ATCATTGGTG TTCAAGATCA AGGAGTTCGC CCCGACA
Vanessa_cardui_NPV367              GAGGAAATCT TACTCGAAGT ATCATTGGTG TTCAAGATCA AGGAGTTCGC CCCCGAC
Wiseana_cervinata_GV342            GAAGAAATAA TGATTGAGGT TTCTCTTCTT TTTAAGATCA AAGAGTTCGC CCCGACA
Wiseana_cervinata_NPV344           GAGGAGGTTC TCTTCGAGGT GTCGCTGGTG TTCAAAATTA AGGAGTTCGC CCCCGAC
Wiseana_signata_NPV                GAGGAGGTTC TCTTCGAGGT GTCGCTGGTG TTCAAAATTA AGGAATTCGC GCCCGAC
XcGV                               GAAGAGGTCA TGATCGAAGT GGCTCTTCTC TTCAAGATTA AGGAATTCGC TCCCGAC
```

***lef-8* Alignment:** In grey are outlined the areas that have been removed from the alignment prior to phylogenetic analyses.

```
10_Lasiocampa_quercus_NPV          NNNNNNGGCA AAAAGTTTTT TCTTtgCCGC AACGCATGTT TGCCCTACGT GGAATACGAA ATGGTTGCGC AACAATTTCA   [80]
106_aporia_crataegi_NPV            NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN ACAAGTTCAA   [80]
112_Abraxas_grossulariata_NPV      GTGGCGGGCA AAAAATTTTT TCTGTGCCGC AACGCGCGGC TGCCCAACGT CGACTACGAG ATGGTGGCCG ATAAGTTTGC   [80]
126_Spilosoma_lutea_NPV            GTGGCGGGTA AAAAGTTTTT TTTGTGCCGC GACACCGTTT TACCGGACGT TGATTACAAA TGCATCGGTG AATTGCTTCA   [80]
14_Anomis_sabulifera_NPV           GTGGCTGGCA AAAAGTTTTT CCTATGTCGC GATGTGCGAT TGCCCAATGT CGACTACGAA ATGGTCGCCA CCAAAATTAA   [80]
174_Erranis_defolaria_NPV          GTGGCAGGTA AAAAGTTTTT TTTGTGCAGA AACGTGAAAC TGCCTAGTGT CGATTACGAA TCGATTAGCA AAAAATTTAA   [80]
175_Anomala_insulata_NPV           GTAGCAGGCA AGAAATTTTT CTTGTGCCGC AACGCGCGGC TGCCCAACGT TGACTACAAT CTAGTAGCCG ACAAATTTCA   [80]
19_Junonia_coenia_GV               GTTGCCGGGA AAAAATTCTT TATATGCAAA GACGTCGATC TGCCCGACCT AAACTACGGA GCGGTCGCGC TGCGCATCCG   [80]
193_Lymantria_obfuscata_NPV        NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN   [80]
195_Trabala_vishnou_NPV            GTGGCTGGCA AAAAATTTTT CTTGTGCCGC AACGTGCGTT TGCCCACGGT CGATTATGAA ATCGTGCGCG ACAAGTTTCT   [80]
202_Lasiocampa_quercus_NPV         GTGGCGGGGA AAAAGTTTTT CTTGTGTCGC AACGTGCGCG TGCCGTCCAT CGATTACGAT CTCGTCAAAT CTAAATTTGA   [80]
205_Kotochalia_NPV                 GTgGCGGGCA AAAAGTTTTT TTTGTGTCGA GACGTCCGTC TCCCGAGTGT GGATTATCCC GCGGTGGCGG TTCCCCTGAG   [80]
228_Hyphantria_cunea_GV            GTGGCGGGAA AAAAATTTTT TTTGTGCCAC AACACCACGT TACCTGACGT GAATTACAAA CATGTTAACG AGCTGTTTCA   [80]
248_Euplasia_lucipara_GV           GTGGCGGGAA AAAAATTCTT CTTGTGTCAC AACGTTGTAT TGCCCGACGT GAACTATGGC CGCGTGTCTG ACGCGTTTAA   [80]
254_Natada_nararia_GV              GTTGCGGGTA AAAAATTTTT TCTTTGTCAT AATACCGTTT TACCCGATGT TAATTATAAA CAAATTAGTA ACATATTTAA   [80]
26_Melanchra_persicae_GV           GTAGCTGGTA AGAAGTTTTT TTTGTGTCAC AATGTTGTTT TACCCGATAT AAACTATAGT CTTGTTAGTG CTATATTCCA   [80]
274_Mamestra_brassicae_NPV         GTGGCGGGCA AAAAGTTTTT CCTGTGCCGT GATGTTCGTT TGCCCAGCGT TGATTATGGC CTAGTAGCTG AAAAATTTAC   [80]
28_Samia_ricini_NPV                GTGGCGGGCA AAAAGTTTTT CCTTTgCCGC AACGCATGTT TGCCCAACGT GGAATACGAA ATGGTTGCGC AAAAATTTCA   [80]
291_O._brumata_NPV_dartmoor        GTTGCGGGCA AGAAATTTTT CTTGTGTCGC GACGTACGAC TGCCCACTAT GAATTATCAA TTGGTCGCGA ATAAATTTGT   [80]
30_Antherea_pernyi_NPV             GTGGCGGGCA AAAAGTTTTT CCTTTGCCGC AACGCATGTT TATCCAACGT GGACTACGAA ATGGTTGCGC AAAAATTCCA   [80]
302lef8c                           NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN   [80]
32_Kotochalia_NPV                  GTgGCGGGCA AAAAGTTTTT TTTGTGTCGA GACGTCCGTC TCCCGAGTGT GGATTATCCC TCGGTGGCGG TTCCCCCTGA   [80]
342_Wiseana_GV                     NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNAT   [80]
344_wiseana_cervinata_NPV          NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNCGA CGACTACNAG GGCGTCGCTC GAAAGTTGCG   [80]
353_Nymphalis_io_NPV               GTAGCGGGCA AAAAATTCTT TTTATGTCGA AACGTGAGGT TGCCGAGCGT AAATTACATG ATAATCGCAG CCAAATTTAA   [80]
362_Shinx_ligustri_NPV             GTGGCGGGCA AAAAGTTTTT CCTTTGCCGC AACGCATGTT TGCCCAACGT GGACTACGAA ATGGTTGCGC AAAAATTTCA   [80]
367_Vanessa_cardui_NPV             GTAGCGGGCA AGAAATTCTT TTTATGTCGA AACGTGAGGT TGCCGAGCGT AAATTACATG ATAATCGCAG CCAAACTTAA   [80]
37_cacoecia_murinana_GV            GTAGCGGGTA AaAAATTTTT CTTGTGCCAC AACACTGTTT TACCGATGgT CGAaTATAAA CGCGTCCCAG ACTTGTTTAG   [80]
'380 Polygonia c-album NPV'        GTAGCGGGCA AAAAATTCTT TTTATGTCGA AACGTGAGGT TGCCGAGCGT AAATTACATG ATAATCGCAG CCAAATTTAA   [80]
384_Pieris_brassicae_GV            GTCGCGGGTA AGAAATTTTT TTTATGCCAC AATACCGTGC TTCCTGACGT TAACTACAAA CATATCAGCC AATTGTTTAC   [80]
402_idea_virgularia_NPV            GTGGCGGGAA AAAAGTTTTT TTTATGTCGC AATACGCGCC TGCCCAACGT GGATTACGAC ATGGTGGCAG ACAAGTTTTC   [80]
436_Pachypasa_papyri_NPV           GTGGCAGGAA AAAAATTTTT TCTATGTCGT AACGTTCGGC TACCGTCGAT TGATTATGAA TATGTAGCTG ACAAGTTTAA   [80]
440lef8c                           NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN   [80]
460_Bombyx_mori_NPV                GTGGCGGGCA AAAAGTTTTT CCTTTGCCGC AACGCGTGTT TGCCTAACGT GAACTTTGAA ATGGTTGCGC AAAAATTTCA   [80]
47_Actias_selene_NPV               GTGGCGGGCA AAAAATTTTT TTTGTGCCGC AACGCGCGGC TGCCCAACGT GGATTACGAC CTGGTGGCTA GCAAATTCCA   [80]
505_buzura_edwardsiiNPV            GTGGCCGGCA AGAAATTTTT TCTGTGCCGC GATGTTAAAC TGCCCAGTGT CGATTATGAA TCTGTAGCTG AAAAGTTTAA   [80]
54_Vanessa_atalanta_NPV            GTAGCGGGCA AAAAATTCTT TTTATGTCGA AACGTGAGGT TGCCGAGCGT AAATTACATG ATAATCGCAG CCAAATTTAA   [80]
545_Darna_trima_GV                 GTAGCTGGCA AAAAATTTTT TTTATGCCAC AAAGCTATTT TACCAGACGT CGATTACAAA TTGATTAGCT CCATATTTAA   [80]
595lef8c                           NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN   [80]
63_Aroa_discalis_NPV               GTGGCGGGCA AAAAGTTCTT TTTGTGTCGC AACGTGCGTC TGCCCCCCGT CAACTATGAC GCGGTGAAAA CGAAATTTAA   [80]
64_Achaea_faber_NPV                GTGGCCGGAA AAAAATTCTT TTTGTGTCGC GACGTACGCT TGCCCAGCGT GGATTATTAT TCGATAGCGT CCAAATTTAA   [80]
66_sp_littoralis_GV                GTGGCGGGAA AGAAATTTTT TCTATGCCAA AATGTAGTTt TGCCAAATCT CCAATACAAA AGAGTTGGGT CTTTAGTGCA   [80]
67_Orgya_mixta_NPV                 GTGGCGGGCA AAAAGTTCTT TTTGTGTCGC AACGTGCGTC TGCCCCCCGT CAACTATGAC GCGGTGAAAA CGAAATTTAA   [80]
692lef8c                           NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNTGGGC TGCCTTCGAT AGATTATAAA TGGTGGCTGA TAAATGCTAA   [80]
744lef8c                           NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN   [80]
750lef8c                           TGGCGGGTTA AAAAATTTTT TCTGTGTCGC AACGTGCGCC TGCCCGCCGG CAATTACCAA GCGGTCGCCG ACAGGTTTGC   [80]
766_Sp_litura_GV                   GTGGCGGGAA AAAAATTTTT TCTATGCAAA AATGTAGTTT TGCCAAATCT TGAATACAAA AGAGTTGGGT CTTTAGTGCA   [80]
768_Euproctis_similis_NPV          GTGGCGGGCA AAAAGTTTTT TCTGTGTCGC AACGCGCGTC TGCCCAACGT GGACTACGAC ACGGTGGCGG CCAAGATGCA   [80]
78_Catopsila_pomona_NPV            GTGGCAGGCA AGAAATTCTT TTTGTGTCGC AACGTGCGTT TGCCCAACGT CGACTACGAA ACGGTCGCCG AACGTGTAAA   [80]
789_Heliconius_erato_NPV           GTGGCGGGCA AAAAATTTTT TCTGTGTCGC AACGCGCGAT TGCCCAACGT GGACTACGAC ATGGTGAGCG ACAAATTTGC   [80]
798_Diacrisia_obliqua_NPV          GTAGCGGGCA AAAAATTTTT TCTGTGTCGC AACGCACGGC TACCTAACGT TGATTATGCT ATGGTGGCCG ACAAGTTTGC   [80]
A1_Mythmnia_unipunctata_NPV        GTGgCgGGCA AAAAGTTTTT CCTCTGCCGC GACGTTCGAT TGCCGAGCGT CGACTATGCG CGAATAGCGT CTAAATTTCA   [80]
a10_Euproctis_chrysorrohea_NPV     GTGGCGGGCA AAAAGTTTTT TCTGTGTCGC AACGCGCGTC TGCCCAACGT GGACTACGAC ACGGTGGCGG CCAAGATGCA   [80]
```

```
a2_Mythmnia_separata_NPV_Japan      GTGGCGGGCA AAAAGTTTTT CCTGTGCCGT GATGTTCGTT TGCCCAGCGT TGATTATGGC CTAGTAGCTG AAAAATTTAC  [80]
a3_Mythmnia_separata_NPV_china      GTGGCGGGCA AAAAGTTTTT CCTGTGTCGC GACGTCCGTC TGCCGAACAT CGATTACCAT GCGGTGGCGA ACAAGTTCAG  [80]
a4_Mythmnia_separata_NPV_India      GTGGCGGGCA AAAAGTTTTT CCTGTGTCGC GACGTCCGTC TGCCGAACGT CGACTACCAT GCGGTGGCGA ACAAGTTCAG  [80]
a5_Cerapteryx_graminis_NPV          GTGGCGGGCA AAAAGTTTTT CCTGTGCCGC GACGTGCGTC TGCCGAGCGT CGACTACGGA TCCGTGGCGG CCAAGTTTAG  [80]
a6_Lacanobia_oleracea_GV            NNNNNNNNNN GGAAGTTTTT CTTGTGTCAC AATGTTGTTT TACCCGATAT AAACTATAGT CTTGTTAGTG CTATATTCCA  [80]
a7_Aglais_urticae_NPV               GTAGCGGGCA AAAAATTCTT TTTATGTCGA AACGTGAGGT TGCCGAGCGT AAATTACATG ATAATCGCAG CCAAATTTAA  [80]
a9_Orgya_antiqua_NPV                GTGGCGGGTA AAAAGTTTTT CTTGTGTCGC AATGTGCGTC TGCCTTCAGT TGATTACGAC GTTGTAAAGT CAAAGTTTGA  [80]
AcMNPV                              GTGGCGGGCA AAAAGTTTTT CCTTTGCCGC AACGCATGTT TGCCCAACGT GGACTACGAA ATGGTTGCGC AAAAATTTCA  [80]
B10_Anticarsia_gemnitalis           GTGGGGGGCA AAAAATTTTT TTTGTGCCGC AACGCGCGCC TGCCTAGCGT TGATTACGAC ATGGTAGCCA ACAAGTTTAC  [80]
b3_Plodia_interpunctella_GV         GTGGCGGGGa AGAAgTTCTT TTTGTGTCAC AACGTGCTAC TGCCCGACGT GAATTACAAG TTCGTAGCCG ACCTATTCAG  [80]
b4_Lymantria_monacha_NPV            GTGGCGGGCA AGAAGTTTTT TCTGTGCCGC GACGTGCGCT TGCCCGGCGT CGACTACCAA TTGGTGGCGG ACAAGTTCTT  [80]
b6_Orgya_leusostigma_NPV            GTGGCCGGTA AGAAATTTTT CTTGTGTCGC AACGTTCGTC TTCCGCCTGT CGACTACGAC GCCATCAAAA CCAAATTTAA  [80]
BmNPV                               GTGGCGGGCA AAAAGTTTTT CCTCTGCCGC AACGCGTGTT TGCCCAACGT GAACTTTGAA ATGGTTGCGC AAAAATTTCA  [80]
c2_Trichoplusia_ni_NPV              GTGGCGGGCA AAAAGTTTTT CCTTTGCCGC AACGCATGTT TGCCCAACGT GGACTACGAA ATGGTTGCGC AAAAATTTCA  [80]
c3_Potato_tuber_GV_8                GTGGCGGGTA AAAAATTTTT TTTGTGTCAC AACACTTTGT TGCCTAATGT CGATTACAAA CTGTCTGCTC AATTGTTTAC  [80]
c4__Potato_tuber_GV_9               GTGGCGGGTA AAAAATTTTT TTTGTGTCAC AACACTTTGT TGCCTAATGT CGATTACAAA CTGTCTGCTC AATTGTTTAC  [80]
C5_Ephasia_cautella_NPV             GTTGCGGGGA AAAAGTTTTT CCTATGCAGA AACGCTCGGT TGCCTAATGT GGATTATCAA TCGGTTGCCA CGAAATTTCA  [80]
CpGV                                GTGGCCGGAA AAAAGTTCTT TTTGTGCCAC AATGTAGTGT TGCCCGATGT GGACTACAAG CGTGTCAGCC TGCTGTTCCG  [80]
242_pieris_napi_GV                  GTAGCGGGTA AGAAATTTTT TTTATGCCAC AATACCGTGC TTCCTGACGT TAACTACAAA CATATCAGCC AATTGTTTAC  [80]
95_pieris_rapae_GV                  GTCGCGGGTA AGAAATTTTT TTTATGCCAC AATACCGTGC TTCCTGACGT TAACTACAAA CATATCAGCC AATTGTTTAC  [80]
CuniNPVlef8                         CGTGCTGGTA AAAACTTTAA CCTCACCACC AACGTTCGGC CAACCTTTGT GGACTTTTGG GGCCTGGTGC GCCTGCTGGA  [80]
EppoNPV                             GTGGCAGGCA AAAAGTTTTT CCTGTGCCGC AACGCGCGTC TGCCCAGTGT CGATTACAAT CTAGTTGCTG ACAAATTTAC  [80]
HaSNPV                              GTGGCTGGAA AAAAATTTTT TTTGTGTCGC GACGTCCGTT TGCCGAGCGT CGATTACATC ATGGTCGCCG AACGTTTTCG  [80]
HzSNPV                              GTGGCTGGAA AAAAATTTTT TTTGTGTCGC GACGTCCGTT TGCCGAGCGT CGATTACATC ATGGTCGCCG AACGTTTTCG  [80]
i3_Panolis_flammea_NPV              GTGGCGGGCA AAAAGTTTTT CCTGTGCCGC GACGTTCGTT TGCCCAGCGT TGATTACGGA TTGGTAGCCG AAAAATTTAA  [80]
K10_Sp_frugiperda_NPV               GTGGCCGGCA AGAAATTTTT TCTGTGCCGC GATGTTAAAC TGCCCAGTGT CGATTATGAA TCTGTAGCTG AAAAGTTTAA  [80]
K11Spodoptera_exempta_NPV           GTGGCCGGAA AAAAATTCTT TTTGTGTCGC GACGTACGCT TGCCCAGCGT GGATTATTAT TCGATAGCGT CCAAATTTAA  [80]
K12_Malacosoma_neustria_NPV         GTGGCGGGTA AAAAATTCTT TTTGTGCCGA AACGTGCGCC TGCCGTCGGT CGACTACGAT TTGGTGCGTG CCAAGTTTGA  [80]
K1lef8c                             GTGGCGGGCA AAAAGTTTTT CCTTTGCCGC AACGCATGTT TGCCCAACGG GGAATATgAA ATGGTTGCGC AACAATTTCA  [80]
K2lef8c                             GTGGAGGGCA AAAAGTTTTT TCTTTGTCGC AACGCATGTT TGCCCAACGT GGATTAtGAA ATGGTTGCGC AAAAATTTAA  [80]
K3lef8c                             GTGAAGGGTA AAAAATTCTT TTTGTGCCGA AACGTGCGCC TGCCGTCGGT CGAATACGAT TTGGTGCGTG CCAAGTTTGA  [80]
K8_sp_sunia_NPV                     GTGGCCGGAA AAAAGTTTTT TCTTTGCCGA AACGTGCGCC TTCCCTCCGT CGATTACCAG AAGGTCGCGG ATAAATTTAA  [80]
K9lef8c                             GTGGTGGGCA AAAAGTTTTT CCTGTGCCGT AATGTTCGTT TGCCCAGCGT TGATTATGGC GTAGAGCTGA AAAATTTCCC  [80]
LdMNPV                              GTGGCGGGCA AAAAGTTTTT CCTGTGTCGC AACGTGCGTC TGCCCGCCGT CGATTACCAA GCGGTCGCCG ACAGGTTTGC  [80]
OpMNPV                              GTGGCGGGGA AAAAGTTCTT TTTATGTCGC AACGCGCGGC TGCCCAACGT CGATTACGCC ATGGTGGCCG ACAAATTTGC  [80]
PxGV                                GTGGCTGGTA AAAAGTTTTT CGTTTGCCAC AATGTTGTGC TAATAGACGT AGATTACACG CTGGTCAGCG ACCTATTTAA  [80]
SeMNPV                              GTGGCGGGCA AAAAGTTTTT CCTCTGTCGC GACGTAAGGC TGCCCAGCGT AGATTACAAT TCAGTGGCCG CGAAATTCAA  [80]
Sp_littoralis_NPV                   GTGGCGGGCA AAAAGTTTTT TCTGTGTCGA GACGTGCGAC TTCCGAACGT CGACTACGAG TCTATAGCGA AAAAATTTAA  [80]
Splitura_NPVgenome                  GTGGCGGGCA AAAAGTTTTT CTTGTGTCGA GACGTGCGGC TTCCAAACGT CGATTACGAA TCTATAGCGA AAAAATTTAA  [80]
XcGV                                GTAGCGGGTA AGAAATTTTT TTTATGCAAC AATGTGGTGC TTCCCAATTT AAATTATAAA CTGGTTGGTG AAAAATTTAG  [80]
M2lef8c                             GTTGCGGGTA AAAAATTTTT TTTGTGTCAT CGTGTAACAT TACCTAATGT CGATTACAAA CGTGTAAGTG ATTTGTTCAA  [80]
N1lef8c                             NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNG CACAATTTAT  [80]
'P1lef8c=E1110'                     NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNTTAA  [80]
CfMNPVlef8.txt                      GTGGCGGGCA AAAAGTTTTT TTTGTGCCGC AACGCGCGGC TGCCCAACGT GGACTACGAC ATGGTAGCCG AAAAGTTTGC  [80]
Nele_NPV                            GCCGCCGGTA AAACGTTCAC AAAAACACAT GATGTCATTT TACCTAATAT AGATTATTCA ACGATATTTC GAATATTTAC  [80]

10_Lasiocampa_quercus_NPV           ATATTTGCTA AAACACAATC TGGTA[illegible]  [illegible]
106_aporia_crataegi_NPV             TCAATTGCTG CAAAACAATT TGATA[illegible]  [illegible]
112_Abraxas_grossulariata_NPV       CGAACTGCAA CAGTTTGGGC TGGTC[illegible]  [illegible]
126_Spilosoma_lutea_NPV             ATTGTTGGTG ACCAAAAAAT TAATT[illegible]  [illegible]
14_Anomis_sabulifera_NPV            ATATTTATTG GAGCAACAGA TATTT[illegible]  [illegible]
174_Erranis_defolaria_NPV           ACAATTGGAG ACGGAAAGTT TAATT[illegible]  [illegible]
175_Anomala_insulata_NPV            CGAGCTGTTG CGCGAGGGCC TGGCA[illegible]  [illegible]
19_Junonia_coenia_GV                CGAACTGATC CGCGACCGAG TGCTG[illegible]  [illegible]
193_Lymantria_obfuscata_NPV         NNNNNNNNNN NNNNNNNNNN NNNNN[illegible]  [illegible]
195_Trabala_vishnou_NPV             CAAGTTGGAG AGCAACATGT TGATT[illegible]  [illegible]
202_Lasiocampa_quercus_NPV          ATATATGGAG AGCAAGCGCC TGATC[illegible]  [illegible]
205_Kotochalia_NPV                  AAAATTGGAG GGAGACGGTG TACTG[illegible]  [illegible]
228_Hyphantria_cunea_GV             CTTATTAATA AATAAAAAAT TGTTA[illegible]  [illegible]
248_Euplasia_lucipara_GV            GAAACTTATC CACCTAAAAC TGTTT[illegible]  [illegible]
254_Natada_nararia_GV               ACAGCTAATA TACAAAGAAC TATTT[illegible]  [illegible]
26_Melanchra_persicae_GV            AAGACTTATA AATGGCAAAT TGTTT[illegible]  [illegible]
274_Mamestra_brassicae_NPV          CGCTTTGCAA AACTCTGGCC TTATT[illegible]  [illegible]
28_Samia_ricini_NPV                 ATATTTGCTA AAACACAATC TGGTC[illegible]  [illegible]
291_O._brumata_NPV_dartmoor         TCATCTCGAG ACAAAGGGTT TGATT[illegible]  [illegible]
30_Antherea_pernyi_NPV              ATATTTGCTA AAACACAATC TGGTC[illegible]  [illegible]
302lef8c                            NNNNNNNNNN NNNNNNNNNN NNNNN[illegible]  [illegible]
32_Kotochalia_NPV                   GAAATTGGAG GGAGACGGTG TACTG[illegible]  [illegible]
342_Wiseana_GV                      AAATTTGATG CCCACGTCGT TGATA[illegible]  [illegible]
344_wiseana_cervinata_NPV           CAAGCTCGTG GCCGAGCGCC TCTTT[illegible]  [illegible]
353_Nymphalis_io_NPV                AAATTTGATC GAGATCGGTC TATTG[illegible]  [illegible]
362_Shinx_ligustri_NPV              ATATTTGCTA AAACACAATC TGGTC[illegible]  [illegible]
367_Vanessa_cardui_NPV              AAATTTGATC GAGGATCGGT CTATT[illegible]  [illegible]
37_cacoecia_murinana_GV             ACATCTCATA AATTCACGAT TGTTA[illegible]  [illegible]
'380 Polygonia c-album NPV'         AAATTTGATC GAGATCGGTC TATTG[illegible]  [illegible]
384_Pieris_brassicae_GV             AAAACTTATT AGTGAAAAAT TATTT[illegible]  [illegible]
402_idea_virgularia_NPV             CGAATTGCGG CAATTTGATC TGGTC[illegible]  [illegible]
436_Pachypasa_papyri_NPV            ATTGTTAGAA GCGCATGATT TATTT[illegible]  [illegible]
440lef8c                            NNNNNNNNNN NNNNNNNNNN NNNNN[illegible]  [illegible]
460_Bombyx_mori_NPV                 ATATTTGCTA AAACACAATC TGGTC[illegible]  [illegible]
47_Actias_selene_NPV                CGAGCTGCTG CGTGCGGGAC TGGCA[illegible]  [illegible]
505_buzura_edwardsiiNPV             ATCTTTGTTG GCTGCCAATC TCATC[illegible]  [illegible]
54_Vanessa_atalanta_NPV             AAATTTGATC GAGATCGGTC TATTG[illegible]  [illegible]
545_Darna_trima_GV                  ACATACAATA GAGATCGGCT TGTTA[illegible]  [illegible]
595lef8c                            NNNNNNNNNN NNNNNNNNNN NNNNN[illegible]  [illegible]
63_Aroa_discalis_NPV                TTGTTTAATC GAGGCGCGCC TCGTG[illegible]  [illegible]
64_Achaea_faber_NPV                 ACACTTGGCC GATCTAAATT TGATC[illegible]  [illegible]
66_sp_littoralis_GV                 ACGCCTAGGA GAATTAGAGC TGTTT[illegible]  [illegible]
67_Orgya_mixta_NPV                  TTGTTTAATC GAGGCGCGCC TCGTG[illegible]  [illegible]
```

692lef8c
744lef8c
750lef8c
766_Sp_litura_GV
768_Euproctis_similis_NPV
78_Catopsila_pomona_NPV
789_Heliconius_erato_NPV
798_Diacrisia_obliqua_NPV
A1_Mythmnia_unipunctata_NPV
a10_Euproctis_chrysorrohea_NPV
a2_Mythmnia_separata_NPV_Japan
a3_Mythmnia_separata_NPV_china
a4_Mythmnia_separata_NPV_India
a5_Cerapteryx_graminis_NPV
a6_Lacanobia_oleracea_GV
a7_Aglais_urticae_NPV
a9_Orgya_antiqua_NPV
AcMNPV
B10_Anticarsia_gemnitalis
b3_Plodia_interpunctella_GV
b4_Lymantria_monacha_NPV
b6_Orgya_leusostigma_NPV
BmNPV
c2_Trichoplusia_ni_NPV
c3_Potato_tuber_GV_8
c4__Potato_tuber_GV_9
C5_Ephasia_cautella_NPV
CpGV
242_pieris_napi_GV
95_pieris_rapae_GV
CuniNPVlef8
EppoNPV
HaSNPV
HzSNPV
i3_Panolis_flammea_NPV
K10_Sp_frugiperda_NPV
K11Spodoptera_exempta_NPV
K12_Malacosoma_neustria_NPV
K1lef8c
K2lef8c
K3lef8c
K8_sp_sunia_NPV
K9lef8c
LdMNPV
OpMNPV
PxGV
SeMNPV
Sp_littoralis_NPV
Splitura_NPVgenome
XcGV
M2lef8c
N1lef8c
'P1lef8c=E1110'
CfMNPVlef8.txt
Nele_NPV

10_Lasiocampa_quercus_NPV
106_aporia_crataegi_NPV
112_Abraxas_grossulariata_NPV
126_Spilosoma_lutea_NPV
14_Anomis_sabulifera_NPV
174_Erranis_defolaria_NPV
175_Anomala_insulata_NPV
19_Junonia_coenia_GV
193_Lymantria_obfuscata_NPV
195_Trabala_vishnou_NPV
202_Lasiocampa_quercus_NPV
205_Kotochalia_NPV
228_Hyphantria_cunea_GV
248_Euplasia_lucipara_GV
254_Natada_nararia_GV
26_Melanchra_persicae_GV
274_Mamestra_brassicae_NPV
28_Samia_ricini_NPV
291_O._brumata_NPV_dartmoor
30_Antherea_pernyi_NPV
302lef8c
32_Kotochalia_NPV
342_Wiseana_GV
344_wiseana_cervinata_NPV
353_Nymphalis_io_NPV
362_Shinx_ligustri_NPV
367_Vanessa_cardui_NPV
37_cacoecia_murinana_GV
'380 Polygonia c-album NPV'
384_Pieris_brassicae_GV
402_idea_virgularia_NPV
436_Pachypasa_papyri_NPV
440lef8c

460_Bombyx_mori_NPV
47_Actias_selene_NPV
505_buzura_edwardsiiNPV
54_Vanessa_atalanta_NPV
545_Darna_trima_GV
595lef8c
63_Aroa_discalis_NPV
64_Achaea_faber_NPV
66_sp_littoralis_GV
67_Orgya_mixta_NPV
692lef8c
744lef8c
750lef8c
766_Sp_litura_GV
768_Euproctis_similis_NPV
78_Catopsila_pomona_NPV
789_Heliconius_erato_NPV
798_Diacrisia_obliqua_NPV
A1_Mythmnia_unipunctata_NPV
a10_Euproctis_chrysorrohea_NPV
a2_Mythmnia_separata_NPV_Japan
a3_Mythmnia_separata_NPV_china
a4_Mythmnia_separata_NPV_India
a5_Cerapteryx_graminis_NPV
a6_Lacanobia_oleracea_GV
a7_Aglais_urticae_NPV
a9_Orgya_antiqua_NPV
AcMNPV
B10_Anticarsia_gemnitalis
b3_Plodia_interpunctella_GV
b4_Lymantria_monacha_NPV
b6_Orgya_leusostigma_NPV
BmNPV
c2_Trichoplusia_ni_NPV
c3_Potato_tuber_GV_8
c4__Potato_tuber_GV_9
C5_Ephasia_cautella_NPV
CpGV
242_pieris_napi_GV
95_pieris_rapae_GV
CuniNPVlef8
EppoNPV
HaSNPV
HzSNPV
i3_Panolis_flammea_NPV
K10_Sp_frugiperda_NPV
K11Spodoptera_exempta_NPV
K12_Malacosoma_neustria_NPV
K1lef8c
K2lef8c
K3lef8c
K8_sp_sunia_NPV
K9lef8c
LdMNPV
OpMNPV
PxGV
SeMNPV
Sp_littoralis_NPV
Splitura_NPVgenome
XcGV
M2lef8c
N1lef8c
'P1lef8c=E1110'
CfMNPVlef8.txt
Nele_NPV

10_Lasiocampa_quercus_NPV
106_aporia_crataegi_NPV
112_Abraxas_grossulariata_NPV
126_Spilosoma_lutea_NPV
14_Anomis_sabulifera_NPV
174_Erranis_defolaria_NPV
175_Anomala_insulata_NPV
19_Junonia_coenia_GV
193_Lymantria_obfuscata_NPV
195_Trabala_vishnou_NPV
202_Lasiocampa_quercus_NPV
205_Kotochalia_NPV
228_Hyphantria_cunea_GV
248_Euplasia_lucipara_GV
254_Natada_nararia_GV
26_Melanchra_persicae_GV
274_Mamestra_brassicae_NPV
28_Samia_ricini_NPV
291_O._brumata_NPV_dartmoor
30_Antherea_pernyi_NPV
302lef8c
32_Kotochalia_NPV
342_Wiseana_GV

```
344_wiseana_cervinata_NPV
353_Nymphalis_io_NPV
362_Shinx_ligustri_NPV
367_Vanessa_cardui_NPV
37_cacoecia_murinana_GV
'380 Polygonia c-album NPV'
384_Pieris_brassicae_GV
402_idea_virgularia_NPV
436_Pachypasa_papyri_NPV
440lef8c
460_Bombyx_mori_NPV
47_Actias_selene_NPV
505_buzura_edwardsiiNPV
54_Vanessa_atalanta_NPV
545_Darna_trima_GV
595lef8c
63_Aroa_discalis_NPV
64_Achaea_faber_NPV
66_sp_littoralis_GV
67_Orgya_mixta_NPV
692lef8c
744lef8c
750lef8c
766_Sp_litura_GV
768_Euproctis_similis_NPV
78_Catopsila_pomona_NPV
789_Heliconius_erato_NPV
798_Diacrisia_obliqua_NPV
A1_Mythmnia_unipunctata_NPV
a10_Euproctis_chrysorrohea_NPV
a2_Mythmnia_separata_NPV_Japan
a3_Mythmnia_separata_NPV_china
a4_Mythmnia_separata_NPV_India
a5_Cerapteryx_graminis_NPV
a6_Lacanobia_oleracea_GV
a7_Aglais_urticae_NPV
a9_Orgya_antiqua_NPV
AcMNPV
B10_Anticarsia_gemnitalis
b3_Plodia_interpunctella_GV
b4_Lymantria_monacha_NPV
b6_Orgya_leusostigma_NPV
BmNPV
c2_Trichoplusia_ni_NPV
c3_Potato_tuber_GV_8
c4__Potato_tuber_GV_9
C5_Ephasia_cautella_NPV
CpGV
242_pieris_napi_GV
95_pieris_rapae_GV
CuniNPVlef8
EppoNPV
HaSNPV
HzSNPV
i3_Panolis_flammea_NPV
K10_Sp_frugiperda_NPV
K11Spodoptera_exempta_NPV
K12_Malacosoma_neustria_NPV
K1lef8c
K2lef8c
K3lef8c
K8_sp_sunia_NPV
K9lef8c
LdMNPV
OpMNPV
PxGV
SeMNPV
Sp_littoralis_NPV
Splitura_NPVgenome
XcGV
M2lef8c
N1lef8c
'P1lef8c=E1110'
CfMNPVlef8.txt
Nele_NPV

10_Lasiocampa_quercus_NPV
106_aporia_crataegi_NPV
112_Abraxas_grossulariata_NPV
126_Spilosoma_lutea_NPV
14_Anomis_sabulifera_NPV
174_Erranis_defolaria_NPV
175_Anomala_insulata_NPV
19_Junonia_coenia_GV
193_Lymantria_obfuscata_NPV
195_Trabala_vishnou_NPV
202_Lasiocampa_quercus_NPV
205_Kotochalia_NPV
228_Hyphantria_cunea_GV
```

```
248_Euplasia_lucipara_GV          TTTTGACGGT GATGCCGAAG TCGAAATGTG TAAAGAAACA CGGAAGCTGT GCAAACAGCC ACGCGACATC AGTTGATACA
254_Natada_nararia_GV             ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
26_Melanchra_persicae_GV          ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
274_Mamestra_brassicae_NPV        ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
28_Samia_ricini_NPV               ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
291_O._brumata_NPV_dartmoor       ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
30_Antherea_pernyi_NPV            ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
302lef8c                          NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
32_Kotochalia_NPV                 ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
342_Wiseana_GV                    ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
344_wiseana_cervinata_NPV         ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
353_Nymphalis_io_NPV              ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
362_Shinx_ligustri_NPV            ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
367_Vanessa_cardui_NPV            ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
37_cacoecia_murinana_GV           ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
'380 Polygonia c-album NPV'       ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
384_Pieris_brassicae_GV           ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
402_idea_virgularia_NPV           ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
436_Pachypasa_papyri_NPV          ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
440lef8c                          NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNATCA CGACGACGTC GACGACGACG
460_Bombyx_mori_NPV               ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
47_Actias_selene_NPV              ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
505_buzura_edwardsiiNPV           ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
54_Vanessa_atalanta_NPV           ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
545_Darna_trima_GV                ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
595lef8c                          NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
63_Aroa_discalis_NPV              ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
64_Achaea_faber_NPV               ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
66_sp_littoralis_GV               ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
67_Orgya_mixta_NPV                ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
692lef8c                          ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
744lef8c                          ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
750lef8c                          ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
766_Sp_litura_GV                  ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
768_Euproctis_similis_NPV         ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
78_Catopsila_pomona_NPV           ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
789_Heliconius_erato_NPV          ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
798_Diacrisia_obliqua_NPV         ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
A1_Mythmnia_unipunctata_NPV       ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
a10_Euproctis_chrysorrohea_NPV    ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
a2_Mythmnia_separata_NPV_Japan    ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
a3_Mythmnia_separata_NPV_china    ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
a4_Mythmnia_separata_NPV_India    ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
a5_Cerapteryx_graminis_NPV        ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
a6_Lacanobia_oleracea_GV          ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
a7_Aglais_urticae_NPV             ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
a9_Orgya_antiqua_NPV              ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
AcMNPV                            ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
B10_Anticarsia_gemnitalis         ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
b3_Plodia_interpunctella_GV       ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
b4_Lymantria_monacha_NPV          ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
b6_Orgya_leusostigma_NPV          ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
BmNPV                             ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
c2_Trichoplusia_ni_NPV            ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
c3_Potato_tuber_GV_8              ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
c4__Potato_tuber_GV_9             ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
C5_Ephasia_cautella_NPV           ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
CpGV                              ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
242_pieris_napi_GV                ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
95_pieris_rapae_GV                ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
CuniNPVlef8                       ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
EppoNPV                           ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
HaSNPV                            ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
HzSNPV                            ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
i3_Panolis_flammea_NPV            ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
K10_Sp_frugiperda_NPV             ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
K11Spodoptera_exempta_NPV         ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
K12_Malacosoma_neustria_NPV       ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
K1lef8c                           ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
K2lef8c                           ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
K3lef8c                           ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
K8_sp_sunia_NPV                   ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
K9lef8c                           ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
LdMNPV                            ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
OpMNPV                            ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
PxGV                              ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
SeMNPV                            ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
Sp_littoralis_NPV                 ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
Splitura_NPVgenome                ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
XcGV                              ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
M2lef8c                           ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
N1lef8c                           ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
'P1lef8c=E1110'                   ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
CfMNPVlef8.txt                    ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
Nele_NPV                          ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------

10_Lasiocampa_quercus_NPV         --TCGTTGCT GATCGCGTTT AATGATAGAC CCACAAATTT AAAATGTTTA AAATCAAACA CGTCGTTTAT CGTGTACACT
106_aporia_crataegi_NPV           --ATGTTAAT GATCGCGTTC AACGATCGTC CGACTAATTT GAAATGTTCA AAACTCGATA TTCCGCGCAT AATATATTAC
112_Abraxas_grossulariata_NPV     --GCGCTGCT GATCGCGTTC AACGACCGGC CCACAAACTT GCAGTGTCTG CGCACCGACG TGCCTAAAAT CGTGTACAGG
```

```
126_Spilosoma_lutea_NPV                --ATACTTTT GGTGGCCTTT AACAATCGAC CCACAATATT TTGGTGCTCC CAATCCGATG TCGTAAAAAT TTACTATCAC
14_Anomis_sabulifera_NPV               --GACATGAT TGTTGCGTTC AACTATCGGC CCACTATTCT CAGATGTCGT AGCGAAGATT TGAAGAGAAT TTTTTACGAA
174_Erranis_defolaria_NPV              --TTACTTCA AATTGCTTTC AACGATAGAC CCACAATATA TTGTTGTCAA CGAAAAGATT TATATAAAAT TTTCTATACA
175_Anomala_insulata_NPV               --ACGTTGCT GATTGCGTTC AACGACCGAC CCACTAACTT GAAATGCCGC CGCGCTGACG TGCCCAAGAT CGTGTACACG
19_Junonia_coenia_GV                   --GGTCTTCT CGTCGCCTTC AACAATCGAC CGACGCGATA CTTCTGTCGT CGCAGCGATT TCGCTCGTAT ATACGAGAAA
193_Lymantria_obfuscata_NPV            NNNNNNNNNN GGTTGCGTTC AACGACCGGC CCACGGTGTA tGCGTGCGGG CGCGCGGACC TCGACAAAAT TTTTTACGCC
195_Trabala_vishnou_NPV                --TTGCTCAG CATCGCTTTC AACGATCGGC CCACGATTTA TTGTTGTTTG CGTCGAGATT TGTACAAAAT CTTTTACGAG
202_Lasiocampa_quercus_NPV             --GTGCTGTT GGTCGCATTC AACAACAGGC CCACGGTTTA CTGTTGCGAC AAAGCCAATC TATCCGTCAT ATTCTATGAG
205_Kotochalia_NPV                     --ACATTTAT TGTCGCCTTC AACAATCGAC CCACCGTTTT CTATTGTCGC AACAGGCATC TCTTGAATGT ATTTTATGAA
228_Hyphantria_cunea_GV                --AAGCTTTT AGTGTCTTTT AACAATAGGC CAACAATATT TTGGTGCACC CAGTCGAATG TTTGCCGAAT TTATTATTAC
248_Euplasia_lucipara_GV               ACACACTTCT CGTGTCTTTT AACAATCGCC CCACCAAGTT CTACTGTAAA CGTGACAATC TTTTGTTTGT GTACTACGAG
254_Natada_nararia_GV                  --AAAATGTT GGTTTCTTTC AACAATAGAC CGACAAAATA TTATTGTTTT AAATCTAATT TATGCGAAAT CTATTATATG
26_Melanchra_persicae_GV               --TGGTTACT GGTGTCGTTT AATAATCGCC CGACCAAGTA CTTTTGTCGA CGCGACAATC TTGTAATAGT ATATTATGAA
274_Mamestra_brassicae_NPV             --GAACTCAT CGTAGCTTTC AATAATCGTC CTACAGTGTA CACGTGTCGC AGAGATCAAC TAGTCTACAT AATCTACGAA
28_Samia_ricini_NPV                    --TCGTTGCT GATCGCGTTT AATGATAGAC CCACAAATTT AAAATGTTTA AAATCAAACA CGTCGTTTAT CGTGTACACT
291_O._brumata_NPV_dartmoor            --ATTTTGCA ATTGTCATTC AACAACAGGC CCACTGTGTA CTCGTGTCGT TATAAAGATT TGACGCGAAT TTTTTACGAG
30_Antherea_pernyi_NPV                 --TCGTTGCT GATCGCGTTT AATGATAGAC CCACAAATTT AAAATGTTTA AAATCAAACA CGTCGTTTAT CGTGTACACT
3021ef8c                               NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN CGCTGTTCAT TGTGTACATT
32_Kotochalia_NPV                      --ACATTTAT TGTCGCCTTC AACAATCGAC CCACCGTTTT TTATTGTCGC AACAGGCATC TCTTGAATGT ATTTTATGAA
342_Wiseana_GV                         --CTCATTAT TATTTCTTTT AACAACCGAC CGACGAAATA CAGGTGCAGT CGAGACGCTC TGTACGCCGT GTACTATGAA
344_wiseana_cervinata_NPV              --GATCTTCT CGTCGCCTTC AACAACCGCC CCACCATCTA CAATTGCAAA AAGACCGATC TTTTGGAGAT CTTTTACGAG
353_Nymphalis_io_NPV                   --AATTTACA AATAGTCTTT AACGACCGAC CGACGATATA TTGGTGTCGC AAAGGAGACT TGTACAAAAT ATTCTATGAA
362_Shinx_ligustri_NPV                 --TCGTTGCT GATCGCGTTT AATGATAGAC CCACAAATTT AAAATGTTTA AAATCAAACA CGTCGTTTAT CGTGTACACT
367_Vanessa_cardui_NPV                 --AATTTACA AATAGTCTTT AACGACCGAC CGACGATATA TTGGTGTCGC AAAGGAGACT TGTACAAAAT ATTCTATGAA
37_cacoecia_murinana_GV                --AACATGCT TGTTGCTTTT AACAATCGTC CGACTAAGTA TTgGTGCGAT AAAAGTAATT TGTACGACAT TTATTACATT
'380 Polygonia c-album NPV'            --AATTTACA AATAGTCTTT AACGACCGAC CGACGATATA TTGTTGTCGC AAAGGAGACT TGTACAAAAT ATTCTATGAA
384_Pieris_brassicae_GV                --AATGTTTT GGTGGCTTTT AACAATCGTC CCACAAAATA TTATTGCGAC AAAAACAATT TGTATTCTAT ATATTATATT
402_idea_virgularia_NPV                --ACGTTATT GATTGCGTTC AATGACAGGC CCACCAATTT GCAGTGCTAc CGCGCCGACG TGCCCAAAAT TGTGTACAGA
436_Pachypasa_papyri_NPV               --ATTTTACC AATAGCCTTT AATAACAGAC CGACTGTTTT TTGTTGTAGA CGATCCGATC TCGGTTACAT TTATTATACG
440lef8c                               AATGGCTAAT GATCGCGTTC AACAATACAC GTACGACGTT TTCGTGCAAG CGAAAGCATC TAGTACGCAT CGTGTACGAG
460_Bombyx_mori_NPV                    --TCGTTGCT GATCGCGTTT AATGATAGAC CCACAAATTT AAAATGTTTA AAATTAAACA CGTCGTTCAT TGTGTACACC
47_Actias_selene_NPV                   --TGGCTGTT GATCGCGTTT AACGACCGGC CCACCAACCT CAAATGCCGC CGCGCTGACG TGCCCAAAAT TGTGTACACG
505_buzura_edwardsiiNPV                --AACGTTTT AATAGCGTTT AATAATCGTC CTACGATCTT TGAATGCGAT AGAGACAAAA TGATCTACAT CATATATCAG
54_Vanessa_atalanta_NPV                --AATTTACA AATAGTCTTT AACGACCGAC CGACGATATA TTGTTGTCGC AAAGGAGACT TGTACAAAAT ATTCTATGAA
545_Darna_trima_GV                     --ACGTTGTT AATAGCTTTT AACAACAGGC CCACCACGTT GTATTGTTCT AGAGACAATT TATGCGCAAT CTACTACAAA
595lef8c                               NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
63_Aroa_discalis_NPV                   --ACTATGCC GGTGGCGTTC AACGACAGGC CCACGGCGTT TTGTTGTCGA CGCGCCGACT TGTTCAAAAT ATTTTACGAG
64_Achaea_faber_NPV                    --AAAGTAAT GGTGGCGTTC AACAACAGAC CCACGATTTA CCAGTGCGAC AAAAGCAATC TACTTTACAT CATTTACGAG
66_sp_littoralis_GV                    --GAAGTGAT GGTTTCGCTC ACCAATAGAC ATATGGTGTT TTGGTGTGAC TGTCAAAATT TGATCAGGAT ATATTATGAG
67_Orgya_mixta_NPV                     --ACTATGCC GGTGGCGTTC AACGACAGGC CCACGGCGTT TTGTTGTCGA CGCGCCGACT TGTTCAAAAT ATTTTACGAG
692lef8c                               --CATATATT GATCGCGTTT AATGATAGAC CTACTATCTA TTATAGTATT AAAGAGAATA TTTTTAGAAT TTTCTATGAA
744lef8c                               --ACTACGCT TGTTGCTTTC AACAATCGTC CGAGTAAATA TTATCGCAAT AAAAGTGATT TGTACGACAT TCATTACATT
750lef8c                               --GCGTGCTG GGTGGCCTTA AACGACCGGC CCACGGAGTA CGTTTGCCGG CGCGCGGATC TCTACAAAAT CTTTTACGCT
766_Sp_litura_GV                       --GAAGTGAT GGTCTCGTTC AACAATAGAC ATACGGTGTT TTGGTGTGAC TGTCAAAATT TGGTCAGGAT ATATTATGAG
768_Euproctis_similis_NPV              --GAACTACT GGTGGCGTTC AACGATCGAC CCACCAACTT GAAATGCCAA AAGAAAAATC TCGCTTTGAT TGTATACTTT
78_Catopsila_pomona_NPV                --ACGCTATT GATCGCGTTC AACGATCGAC CGACAAATTT AAAATGCTCA AAATCCAACG TGTCCGCCAT CGTGTACAAC
789_Heliconius_erato_NPV               --TCGCTCCT AATCGCCTTC AACGACCGGC CCACCAATTT ACAATGCGCG CGCACCGACG TGCCCAAAAT TGTGTACACG
798_Diacrisia_obliqua_NPV              --ACGCTTCT CATCGCTTTT AATGATCGGC CGACCAATTT GCAATGCTTG CGCGCTGATG TACCTAAAAT TGTTTACCGC
A1_Mythmnia_unipunctata_NPV            --CGATTGAT TATCGCCTTC AATAACCGTC CGACGATATA CACGTGTCGT GCCGATCGTC TGCTGTACAT TGTGTACGAA
a10_Euproctis_chrysorrohea_NPV         --GAACTACT GGTGGCGTTC AACGATCGAC CCACCAACTT GAAATGCCAA AAGAAAAATC TCGCTTTGAT TGTATACTTT
a2_Mythmnia_separata_NPV_Japan         --GAACTCAT CGTAGCTTTC AATAATCGTC CTACAGTGTA CACGTGTCGC AGAGATCAAC TAGTCTACAT AATCTACGAA
a3_Mythmnia_separata_NPV_china         --TGGCTGAT GATAGCCTTC AACAACCGGC CCACGATCTA CAAGTGTCGG CGGAAACATT TGCCCTATCT GACGTACGAA
a4_Mythmnia_separata_NPV_India         --TGGCTGAT GATAGCCTTC AACAACCGGC CCACGATCTA CAAGTGTCGG CGGAAACATT TGCCCTATCT GACGTACGAG
a5_Cerapteryx_graminis_NPV             --CGGATCAT GATCGCGTTT AACAATCGGC CCACCATCTA TCGGTGCCGC ATTGAAAAAT TGATCTACAT TATTTACGAA
a6_Lacanobia_oleracea_GV               --TGGTTACT GGTGTCGTTT AATAATCGCC CGACCAAGTA CTTTTGTTGT CGTGACGATC TTGTCACGGT GTATTATGAG
a7_Aglais_urticae_NPV                  --AATTTACA AATAGTCTTT AACGACCGAC CGACGATATA TTGTTGTCGC AAAGGAGACT TGTACAAAAT ATTCTATGAA
a9_Orgya_antiqua_NPV                   --TTGCTGCG CGTGGCTTTC AACGACCGCC CAACGGTGTT TTGTTGCAAG CGCGCAGATT TGTATAAAAT ATTTTACGAG
AcMNPV                                 --TCGTTGCT GATCGCGTTT AATGATAGAC CCACAAATTT AAAATGTTTA AAATCAAACA CGTCGTTTAT CGTGTACACT
B10_Anticarsia_gemnitalis              --CTGTTGCT TATTGCGTTC AACGACAGAC CCACCAATTT AAAATGCCGC CGGGCCAACG TGCCCGAAAT AGTTTACGCT
b3_Plodia_interpunctella_GV            --ATCATGAT GGTATCATTT AATAACCGAC CGACCAAATT CTATTGCGAC TCACAGCATA TTCTGGATGT ATACTATCAT
b4_Lymantria_monacha_NPV               --TATTACAG GATCGCGTTC AACGACCGGC CCACCGTGTT CGCGTGCCAA CGAAAAAACC TCGACAAATT ATTTTACATC
b6_Orgya_leusostigma_NPV               --ATATTGCA CGTGGCGTTC AACGATCGCC CGACGGTGTT TTGTTGTCGT CGCACTGATC TACGCAAGAT ATTTTACGAG
BmNPV                                  --TCGTTGCT GATCGCGTTT AATGATAGAC CCACAAATTT AAAATGTTTA AAATTAAACA CGTCGTTCAT TGTGTACACT
c2_Trichoplusia_ni_NPV                 --TCGTTGCT GATCGCGTTT AATGATAGAC CCACAAATTT AAAATGTTTA AAATCAAACA CGTCGTTTAT CGTGTACACT
c3_Potato_tuber_GV_8                   --GATTTGAT GGTAGCTTTC AACAATCGTC CCACCAAGTT TTATTGTAAT CGAGACAATT TGTACACGGT ATATTACACT
c4__Potato_tuber_GV_9                  --GATTTGAT GGTAGCTTTC AACAATCGTC CCACCAAGTT TTATTGTAAT CGAGACAATT TGTACACGGT ATATTACACT
C5_Ephasia_cautella_NPV                --ATGTTGCT CGTCGCTTTC AACGACCGCC CTACGAATCT CAAGTGCAAC ATTTCCGATC TATCAAAAAT CATATACGAA
CpGV                                   --GAACTGGT GGTGGCGTTT AATAATCGTC CCACCAAGTA CTATTGCAAC AGGGATTCCT TGTACCAAAT CTACTACACT
242_pieris_napi_GV                     --AATGTTTT GGTGGCTTTT AACAATCGTC CCACAAAATA TTATTGCGAC AAAAACAATT TGTATTCTAT ATATTATATT
95_pieris_rapae_GV                     --AATGTTTT GGGGGCTTTT AACAATCGTC CCACAAAATA TTATTGCGAC AAAAACAATC TGTATTGTAT ATATTATATT
CuniNPVlef8                            --GAGCTGTA CGTTTACCTA AACACCATCC CCACCCATCT ACGGGCAACC GAGACCAAGT TGATACCGCT GTTTGAGGCC
EppoNPV                                --ACGTTGTT GATCGCGTTT AACGACAGAC CGACAAATTT GCAATGTCGA CGAACAGACG TGCCTACTAT CGTGTACACA
HaSNPV                                 --TTACTCTT GATCGCATTT AACGATCGTC CCACAGTGTT CAATTGTCGA CGAACTGATT TACAAACGAT TTTCTATTAT
HzSNPV                                 --TTACTCTT GATCGCATTC AACGATCGTC CCACAGTGTT CAATTGTCGA CGAACTGATT TACAAACGAT TTTCTATTAT
i3_Panolis_flammea_NPV                 --GAACTCAT CGTGGCTTTC AATAATCGTC CTACAGTGTA CACGTGTCGC AGAGATCAAC TAGTCTACAT AATCTACGAA
K10_Sp_frugiperda_NPV                  --AACGTTTT AATAGCGTTT AATAATCGTC CTACGATCTT TGAATGCGAT AGAGACAAAA TGATCTACAT CATATATCAG
K11Spodoptera_exempta_NPV              --AAAGTAAT GGTGGCGTTC AACAACAGAC CCACGATTTA CCAGTGCGAC AAAAGCAATC TACTTTACAT CATTTACGAG
K12_Malacosoma_neustria_NPV            --ACGATACT GATCGCCTTC AACAATAGAC CTACGGTGTT CTGCTGCCAC ACGTCCAGTC TGACCATTTT ATTTTACGAA
K1lef8c                                --TTGTTGTT GATTGCGTTT AATGATAGAC CCACAAATTT AAAATGTTTA AAATCAAACA CGTCGTTTAT CGTGTACACT
K2lef8c                                --TCGTTGTT GATTGCGTTT AATGATAGAC CCACAAATTT AAAATGTTTA AAATCAAACA CGTCGTTTAT CGTGTACACC
K3lef8c                                --GCGATACT GATCGCCTTC AACAATAGAC CTACGGTGTT CTGCTGCCAC ACGTCCAGTC TGACCATTTT ATTTTAcGAA
K8_sp_sunia_NPV                        --AATATCAG AATAGCGTTC AACGACCGAC CGACTATATA CGCGTGTGGT TTCGACAATT TGCAAATAAT TTTCTATCAT
K9lef8c                                --GAACTCAT CGTAGCTTTC AATAATCGTC CTCCAGTGTA CACGTGTCGC AGAGATCAAC TAGTCTACAT AATCTACGAA
LdMNPV                                 --GAGCTCCG CGTCGCGTTC AACGACCGGC CCACGGTGTA CGCGTGCCGG CGCGCGGACC TCGACAAAAT TTTTTACGCC
OpMNPV                                 --GCCTTGCT GATCGCGTTC AACGACCGCC CCACCAACTT GCAGTGCGCG CGCGCCGACG TGCCCAAAAT TGTGTACCGG
PxGV                                   --GTATATTT AATTTCTTTC AACAACAGAC CTACTAATTT GTATTGCGAT AAAGACGATG TTTGTAAAAT TTACTATCTA
SeMNPV                                 --AATGTGTT GATTGCATTT AACAATCGTC CGACCATCTT TCAATGCGAT CGCAACAAGC TCGTATACAT TGTGTACGAA
Sp_littoralis_NPV                      --TGGCTAAT GATCGCGTTC AACAATAGAC CTACGACGTT TTCGTGCAAG CGAAAGCATC TAGTACGCAT CGTGTACGAG
Splitura_NPVgenome                     --TGGTTGAT GATCGCTTTC AATAATAGAC CTACGACGTT TTCGTGCAAA CGAAAGCATC TAGTGCGCAT CGTGTACGAG
XcGV                                   --GTGGTTAT GGTGTCTTTC AACAATAGAC CGACCGTGTT AAAATGTAAG CCCGACGATT TAATTCCCAT TTATTATGAA
```

```
M2lef8c                        ---ATTTTACT TGTGAGCTAT AATAACAGAC CTACTAAATA TTTTTGTAGA AAAGATTCGT CATTCTATAC AGTTTATTAT
N1lef8c                        ---ACGCTTCT CATCGCCTTT AATGACCGGC CAACCAATTT GCAATGCTTG CGCGCCGATG TGCCTAAAAT TGTTTACCGC
'P1lef8c=E1110'                ---NATATACT GGTAGCGTTC AACAATCGAC CCACCAAATT CATGTGTAAA CGGGCAAAAT CTCTACATAA TCTACCATCG
CfMNPVlef8.txt                 ---TCGTTGCT CGTCGCGTTC AACGATCGGC CCACGAATTT GCAGTGCTCG CTCGCCGACG TGCCAAAAAT CCTGTACAGG
Nele_NPV                       ---AAATATCG AATCGTTTTT AATTCTTGTC CAACACAATA T------TTT GTACACAATG TGGAAAAATT TGTA------

10_Lasiocampa_quercus_NPV      ---ATGAAAA GAAACATGGC GCCCATCGAA TTGAAGATTA CAGACAGA-- ----ATACTT TACGTTAATC ATCACGAAGG
106_aporia_crataegi_NPV        ---ATTAAAC GCAACATGTG TCCGATCGAA TTGAAATATT CAAAGGAG-- ----ATTTTG TACATTAACC ATCACGAAGG
112_Abraxas_grossulariata_NPV  ---CTGAAGC GCGACCGTTG CCCCATCGAG CTGAAGCTGT CGCCGCGA-- ----ATTTTG CACGTGAACC ACCACGAGGG
126_Spilosoma_lutea_NPV        ---ATCAAAC GAAACGCCTC GCCCGTTGAG GTCAAAATGA ACGACCAC-- ----ATACTT TTTATAAACC ACCACGAAGG
14_Anomis_sabulifera_NPV       ---TTCAAGC GACAATTATG TCCGGTAGAA CTTCGTTTGG GCAAAGGC-- ----ATTTTG TTCATTAACC ATCATGAAGG
174_Erranis_defolaria_NPV      ---TTTAATC ACAACATGTA TCCTGTCGAA TTGAAACTGT CCAATAAC-- ----ATTTTA TTGGTAAACC ATCACGAAGG
175_Anomala_insulata_NPV       ---CTCAAAC GAAACCGATG CCCAATTGAG TTGAAGGCGT CGCACAGC-- ----ATTTTG TACGTGAACC ACCACGAGGG
19_Junonia_coenia_GV           ---CTGAAAA GCGATCGGTC GCCGGTCGAA TTGAAACTCG CCGACGAC-- ----GTTCTA TTCATAAATC ATCACGAAGG
193_Lymantria_obfuscata_NPV    ---TTCAAGC GCAACATGTC GCCGATCGAG CTGAAACTCA AGGACGGC-- ----ATACTG TTCGTGAACC ACCACGAGGG
195_Trabala_vishnou_NPV        ---TTTAAGC GACAGAATTG CGCCATAGAA CTCAAATTGT GCAACGGC-- ----ATTCTG TTCGTCAACC ATCACGAGGG
202_Lasiocampa_quercus_NPV     ---CTGAAGC GCAATATGTG CCCCATAGAG GTGAAGTTGT TTGACGAC-- ----ATTTTA TTCATAAATC ACCATGAAGG
205_Kotochalia_NPV             ---TTAAAAC ACAATATGAG TCCCGTTGAA ATGAAATTGG CGAACGGA-- ----ATTCTC TTTATCAATC ATCACGAGGG
228_Hyphantria_cunea_GV        ---GTTAAAC GAAACATGTC CCCAGTTGAG GTTAAAATGA ATGACAAC-- ----ACATTA TTTTTAAACC ACCACGAAGG
248_Euplasia_lucipara_GV       ---TTGAAAC GAAACTACGC TCCAATAGAG ATAAAGGTGT CCAACAAC-- ----ATATTA TTTGTAAATC ATCACGAGGG
254_Natada_nararia_GV          ---TTAAAAA GATCTTATAC ACCTGTTGAA ATAAAAGTGG CTAACAAT-- ----ATTTTG TTTGTAAATC ACCACGAAGG
26_Melanchra_persicae_GV       ---TTAAAGA GATCTTTTAG TCCGATAGAA GTAAAAGTTT CAGACAAT-- ----CTTTTG ATGGTTAATC ATCACGAAGG
274_Mamestra_brassicae_NPV     ---CTAAAAC GAAATCGCTT TCCCATTGAA TTGAAACTCA CCGACAAA-- ----ATACTG TTTATCAACC ATCACGAAGG
28_Samia_ricini_NPV            ---ATGAAAA GAAACATGGC GCCCATCGAA TTGAAGATTA CAGACAGA-- ----ATACtT TTCGTTAATC ATCACGAAGG
291_O._brumata_NPV_dartmoor    ---TTGAAAC ACGATCAGTG CCCGATAGAA TTAAAACTAA CCGATACG-- ----ACGCTG TACGTCAACC ATCACGAAGG
30_Antherea_pernyi_NPV         ---ATGAAAA GAAACATGGC GCCCATCGAA TTGAAGATTA CAGACAGA-- ----ATACTT TACGTTAATC ATCACGAAGG
302lef8c                       ---TTAAAGG GTCCCATGTC GCCTATAAAA TTGAAAATTA CCGACAAA-- ----ATACTG TAcGTGAATC ACCATGAAGG
32_Kotochalia_NPV              ---TTAAAAC ACAATATGAG TCCCGTTGAA ATGAAATTGG CGAACGGA-- ----ATTCTC TTTATCAATC ATCACGAGGG
342_Wiseana_GV                 ---ATGAAGC GCAATTTTAA GCCGATAGAA ATAAAACCAG CCAACAAA-- ----ATACTA TTTATCAACC ACCACGAAGG
344_wiseana_cervinata_NPV      ---CTGAAGC ACGACCGCGC GCCCATCGAG GTGAAACTGT CGGGCGAC-- ----ATACTT TTTCTCAACC ACCACGAGGG
353_Nymphalis_io_NPV           ---CTAAAAC GTAATTCGTG TCCGATTGAA ATAAAAATCG CCAAACGA-- ----TTTCTG TTCATTAATC ATCACGAAGG
362_Shinx_ligustri_NPV         ---ATGAAAA GAAACATGGC GCCCATCGAA TTGAAGATTA CAGACAGA-- ----ATACTT TACGTTAATC ATCACGAAGG
367_Vanessa_cardui_NPV         ---CTAAAAC GTAATTCGTG TCCGATTGAA ATAAAAATCG CCAAACGA-- ----TTTCTG TTCATTAATC ATCACGAAGG
37_cacoecia_murinana_GV        ---CTGAAAA GAAATCGTTC TCCGATCCAA ATAAAAGTGG CTAACAAC-- ----ATATTG TTTGTAAACC ATCACGAAGG
'380 Polygonia c-album NPV'    ---CTAAAAC GTAATTCGTG TCCGATTGAA ATAAAAATCG CCAAACGA-- ----TTTCTG TTCATTAATC ATCACGAAGG
384_Pieris_brassicae_GV        ---TTAAAAC GTTCCAATAC TCCTGTTGAA ATAAAATTTA TTGATAAT-- ----ATATTA TTTATTAACC ACCACGAAGG
402_idea_virgularia_NPV        ---TTGAAGC GCGACCGTTG CCCCATCGAA CTTAAAGTTT CACCGAGT-- ----ATTTTG TATGTAAACC ACCACGAAGG
436_Pachypasa_papyri_NPV       ---CTAAAAC ACAACATGTC TCCCATAGAA ATAAAATTAA TAGACGAT-- ----ATGTTG TTCGTAAACC ACCATGAAGG
440lef8c                       ---TTTAAAC GCTAACGGTT TCCAGTCGAG ATCAAACTGT CCAAAAGT-- ----ATTCTA TTTGTAAATC ATCACGAAGG
460_Bombyx_mori_NPV            ---ATGAAAA GAAACATGGC GCCCATCGAA TTGAAGATTA CAGACAGA-- ----ATACTT TACGTTAATC ATCACGAAGG
47_Actias_selene_NPV           ---CTCAAGA TCAACCGGTG CCCGATCGAG CTTAAAGCGT CGCGAAGC-- ----ATTCTG TACGTGAACC ATCACGAAGG
505_buzura_edwardsiiNPV        ---TTGAAAC GTAACAAGTT TCCCATTGAA ATTAAATTTC ACAACAAC-- ----ATTTTG TTTCTGAACC ATCACGAAGG
54_Vanessa_atalanta_NPV        ---CTAAAAC GTAATTCGTG TCCGATTGAA ATAAAAATCG CCAAACGA-- ----TTTCTG TTCATTAATC ATCACGAAGG
545_Darna_trima_GV             ---TTTAAAC ACAACAATTC TCCTATAGAA ATAAAATTAC ACAACAAT-- ----ATCCTA TTTATTAACC ATCACGAAGG
595lef8c                       ---NNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNAATC ATCATGAAGG
63_Aroa_discalis_NPV           ---TTTAAGC GCAACATGTG TGCGGTGGAA TTGAAAGTGG CCAACAAT-- ----ATTTTG TTTGTCAACC ACCACGAAGG
64_Achaea_faber_NPV            ---TTTAAAC GAAACCAGTT TCCTGTAGAG CTAAAATTTC ACGACAAC-- ----ATCTTG TTCGTCAATC ATCACGAAGG
66_sp_littoralis_GV            ---CTCAAGC GTAGTTACTC ACCTATAGAA ATACAGTATA GTAGTAAG-- ----ATTCTT TTTATTAACC ACCTCGAGGG
67_Orgya_mixta_NPV             ---TTTAAGC GCAACATGTG TGCGGTGGAA TTGAAAGTGG CCAACAAT-- ----ATTTTG TTTGTCAACC ACCACGAAGG
692lef8c                       ---TTGAAAC GTAATTATTG TCCGATAGAA TTAAAATTAG AAGAAAAT-- ----ATACTA TTTATAAATC ATCATGAAGG
744lef8c                       ---TTAAAAA GGAATCGTTC GCCAATAGGA ATAAAAGTGG CTAACAAT-- ----ATATTG TTTGTCAATC ATCACGAAGG
750lef8c                       ---TTCAAAC GCAACATGTc GCCGATTGAG CTGAAACTCA GCAACGAC-- ----ATATTG TTTGTAAACC ACCACCAAGG
766_Sp_litura_GV               ---CTCAAGC GAAACTACTC ACCTATAGAA ATAAAGTGTA GTAATAAT-- ----ATTCTT TTTATTAacC ACCCcGAGGG
768_Euproctis_similis_NPV      ---ATAAAGC GCAATTTGGC ACCGATAGAG TTGCGTTTAA CCGACAGA-- ----ATATTA TTTGTCAACC ATCACGAGGG
78_Catopsila_pomona_NPV        ---ATGAAAC GCGACATGTC GCCGATAGAA TTAAAAATTA GCGACAAA-- ----ATACTT TTTGTCAATC ACCACGAGGG
789_Heliconius_erato_NPV       ---CTGAAAC GCAACCGCTG TCCGATCGAG CTCAAGGCGT CCGCCAAA-- ----GTTTTG TACGTGAACC ACCACGAGGG
798_Diacrisia_obliqua_NPV      ---CTTAAAC GCAACCGTTG CCCTATCGAA CTAAAAGTGT CCCAAAAT-- ----ATTTTG TGTGTAAACC ACCACGAGGG
A1_Mythmnia_unipunctata_NPV    ---TTGAAAC GCAACCGGTT TCCCGTCGAA CTCAAACTCA CCGACAAC-- ----ATTCTC TTCATCAACC ATCACGAAGG
a10_Euproctis_chrysorrohea_NPV ---ATAAAGC GCAATTTGGC ACCGATAGAG TTGCGTTTAA CCGACAGA-- ----ATATTA TTTGTCAACC ATCACGAGGG
a2_Mythmnia_separata_NPV_Japan ---CTAAAAC GAAATCGCTT TCCCATTGAA TTAAAACTCA CCGACAAA-- ----ATACTG TTTATCAACC ATCACGAAGG
a3_Mythmnia_separata_NPV_china ---CTGAAAC GAAACCGTTT CCCCGCCGAG ATCAAGCTGA ACTCGAAC-- ----ATTCTG TTCGTGAACC ATCACGAAGG
a4_Mythmnia_separata_NPV_India ---CTGAAAC GAAACCGTTT CCCCGCCGAG ATCAAGCTGA ACTCGAAC-- ----ATTCTG TTCGTGAACC ATCACGAGGG
a5_Cerapteryx_graminis_NPV     ---CTGAAGC GCAACCGGTT CCCGGTCGAA ATAAAGTTCC ACGACAAC-- ----ATACTC TTCGTCAACC ACCACGAGGG
a6_Lacanobia_oleracea_GV       ---TTAAAGA GATCTTTTAG TCCGATAGAA GTAAAAGTTT CAGACAAT-- ----CTTTTG ATGGTTAATC ATCACGAAGG
a7_Aglais_urticae_NPV          ---CTAAAAC GTAATTCGTG TCCGATTGAA ATAAAAATCG CCAAACGA-- ----TTTCTG TTCATTAATC ATCACGAAGG
a9_Orgya_antiqua_NPV           ---ATTAAAC GCAACGCATG TGCCATCGAG CTGAAATTAA CCGACGAC-- ----ATTTTG TTTGTAAATC ATCACGAAGG
AcMNPV                         ---ATGAAAA GAAACATGGC GCCCATCGAA TTGAAGATTA CAGACAGA-- ----ATACTT TACGTTAATC ATCACGAAGG
B10_Anticarsia_gemnitalis      ---TTTAAAC GTCACCGCTG CCCCATCGAG CTAAAAATGT CGCAAAGC-- ----ATTTTG TACGTTAACC ATCACGAAGG
b3_Plodia_interpunctella_GV    ---TTCAAAC GCAGCTACAG CCCGGTAGAG ATCAAAATCA GCGACGGT-- ----ATCCTC TTCATCAATC ATCACGAGGG
b4_Lymantria_monacha_NPV       ---TTCAAAC GCAACCAATC GCCCGTCGAA CTCAAACTAC AGGGAGAC-- ----ACGCTG TTCGTCAACC ATCACGAGGG
b6_Orgya_leusostigma_NPV       ---TTTAAAC GAAACGCGTG CGCCGTAGAA TTGAAGTTGG CTAACGAT-- ----ATTTTG TTTGTGAATC ACCACGAAGG
BmNPV                          ---ATGAAAA GAAACATGGC GCCCATCGAA TTGAAGATTA CAGACAGA-- ----ATACTT TACGTTAATC ATCACGAAGG
c2_Trichoplusia_ni_NPV         ---ATGAAAA GAAACATGGC GCCCATCGAA TTGAAGATTA CAGACAGA-- ----ATACTT TACGTTAATC ATCACGAAGG
c3_Potato_tuber_GV_8           ---TTAAAAC GGCGCCAATC ACCGATCGAA ATCAAAGTTA GCAACGGC-- ----ATCCTT TTTGTCAATC ATCACGAAGG
c4__Potato_tuber_GV_9          ---TTAAAAC GGCGCCAATC ACCGATCGAA ATCAAAGTTA GCAACGGC-- ----ATCCTT TTTGTCAATC ATCACGAAGG
C5_Ephasia_cautella_NPV        ---ATGAAAC GCAACATGTT TCCTATTGAA ATGAAAACCA ACAATAAA-- ----ATTTTA TTTATAAACC ATCATGAAGG
CpGV                           ---ATGAAGA GGGGATACGC GCCCGTGGAG ATAAAACTCA CCAACAAC-- ----ATTTTG TTTGTGAACC ACCATGAAGG
242_pieris_napi_GV             ---TTAAAAC GTTCCAATAC TCCTGTTGAA ATAAAATTTA TTGATAAT-- ----ATATTA TTTATTAACC ACCACGAAGG
95_pieris_rapae_GV             ---TTAAAAC GTTCCAATAC TCCTGTTGAA ATAAAATTTA TAGATAAT-- ----ATATTA TTTATTAACC ACCACGAAGG
CuniNPVlef8                    ---CTGAAGG CGAACCGGGC CCAGGTTGAG CTCCGCTTCT GGCCCGAACA AAAAGTGCTG CAAATTAACC ACTGCCCGGG
EppoNPV                        ---TTAAAAC GCAATCGGTG TCCAATTGAG TTAAAAATGT CGCAAAGT-- ----ATTCTG TACGTAAACC ATCATGAAGG
HaSNPV                         ---TTCAAAC GAAATATGTG TCCAGTAGAA CTGAAATTAG CCAACAAC-- ----ATTTTG TTTGTCAACC ATCACGAGGG
HzSNPV                         ---TTCAAAC GAAATATGTG TCCAGTAGAA CTGAAATTAG CCAACAAC-- ----ATTTTG TTTGTCAACC ATCACGAGGG
i3_Panolis_flammea_NPV         ---CTGAAAC GTAACCGCTT TCCTATCGAA TTAAAACTGA CCGACAGA-- ----ATACTC TTTATCAACC ATCACGAAGG
K10_Sp_frugiperda_NPV          ---TTGAAAC GTAACAAGTT TCCCATTGAA ATTAAATTTC ACAACAAC-- ----ATTTTG TTTCTGAACC ATCACGAAGG
K11Spodoptera_exempta_NPV      ---TTTAAAC GAAACCAGTT TCCTGTAGAG TTAAAATTTC ACGACAAC-- ----ATCTTG TTCGTCAATC ATCACGAAGG
K12_Malacosoma_neustria_NPV    ---TTGAAAC GCAACTCGTG TCCCATTGAG TTTAAAATCT TCGACAAA-- ----ATTCTT TTCATCAATC ATCACGAGGG
K1lef8c                        ---ATGAAAA GAAACATGGC GCCCATCGAA TTGAAGATTA CAGACAGA-- ----ATACTT TCcGTTAATC ATCACGAAGG
K2lef8c                        ---ATGAAAA GAAACTTGGC ACCCATAGAA TTGAAGATTA CAGTCAGA-- ----ATACTT TCCGTTAATC ATCACGAAGG
```

```
K3lef8c                          ---TTGAAAC GCAACTCGTG TCCCATTGAG TTTAAAATCT TCGACAAA-- ----ATTCTT TTCATCAATC ATCACGAGGG
K8_sp_sunia_NPV                  ---CTGAAAC GGCAATACAG TCCCATCGAA CTGAAGCTGG CCAACCGC-- ----ATTCTT TTCGTTAATC ACCACGAAGG
K9lef8c                          ---CTAAAAC GAAATCGCTT TCCTATTGAA TTGAAACTCA CCGACAAA-- ----ATACTG TTTATCAACC ATCACGAAGG
LdMNPV                           ---TTCAAGC GCAACATGTC GCCGATCGAG CTGAAACTCA ACGACGGC-- ----ATCCTG TTCGTGAACC ACCACGAGGG
OpMNPV                           ---CTAAAGC GCGACCGCTG CCCGATCGAG CTAAAGGCGT CCTCGAAC-- ----ATTTTG TACGTCAACC ACCACGAGGG
PxGV                             ---ATGAAAC GGAATTTTTC GCCCGTCGAA ATAAAACTGA CTAAAGAT-- ----ATTTTA TTCATCAATC ACCACGAGGG
SeMNPV                           ---CTGAAAC GCAACAAGTT TCCCATCGAA ATAAAATTTC ACAACGAC-- ----ATTATT TTTTTGAACC ATCACGAGGG
Sp_littoralis_NPV                ---TTTAAAC GCAAACGGTT TCCAGTCGAG ATCAAACTGT CCAAAAGT-- ----ATTCTA TTTGTCAATC ATCACGAAGG
Splitura_NPVgenome               ---TTTAAAC GCAAACGGTT TCCCGTCGAG ATCAAACTGT TTAAAAAT-- ----ATTCTA TTTGTCAATC ATCACGAAGG
XcGV                             ---TTAAAAC GAGGGTACTC GCCCGTTGAA CTTAAGTACG CCAATGGT-- ----ATTTTG TTCGTAAACC ATCACGAAGG
M2lef8c                          GTTTTAAAAC GCAATTATTC ACCTATCGAA TTAAAATTGT GTGACAAC-- ----ATATTA TTTGTAAATC ATCATGAAGG
N1lef8c                          ---CTTAGAC GCAACCGTTG CCCTATCGAA CTAAAAGTAT CCCAAAAT-- ----ATTTTG TGTGTAAACC ACCACGAGGG
'P1lef8c=E1110'                  --GTTGAAAA ACAACCAAAC CGCTGTCGAA ATGAAATTTA CCAACGAT-- ----ATTCTA TTCATCAACC ATCACGAGGG
CfMNPVlef8.txt                   ---CTGAAGC GCAACCGCTG TCCGATCGAG CTCAAAGTTT CCTTGAAA-- ----ATTTTG TACGTAAACC ACCACGAAGG
Nele_NPV                         ---AAAATAA TCAAGGCATT GCCAATAGAA ATACGTATGT TTAAATCG-- ----TTGATC TGTTTGAATT ACCACGTTGG

10_Lasiocampa_quercus_NPV        AATGATATGC ATCAAAAAAA AGTTGCGGGT GAACAACGAA ---------- -----GCCGA CATTAACGTA TTGTTGACAC
106_aporia_crataegi_NPV          AATGATTTGT TTAAAAAAAG AAATGCTCAT CAATAACACG ---------- -----ACAAA AATAAACACG TTGCTCACTC
112_Abraxas_grossulariata_NPV    CATGGTGTGC ATAGTTAAAA AGGTGCGCAT ACCCCCCCCC CCCCAC---- -----GCCGA TATTTACGCG CTGCTGACCC
126_Spilosoma_lutea_NPV          TATGATTATG TTGCGCAAAG TGGTGTGTAT CAATAAAACC ---------- -----GTTTT TATAAACACC CTGCAAACGC
14_Anomis_sabulifera_NPV         TATGATATGT TTGAAAAAAA AAATTGATCT GAAAACC--- ---------- -----GCAAC GATTAACGCA CTCTTGACCG
174_Erranis_defolaria_NPV        CATGATATGT ATAAAGCGAA GAATAGCAAT AAAACCCCCC GAC------- -----GTAAT CATTAATGCG CTCTTGACTC
175_Anomala_insulata_NPV         CATGGTGTGT GTGGCCAAAA AGCTGCGCAT GCCCCCCCCC CCCCAC---- -----GTCAC AGTTTACGCG CTGCTCACGC
19_Junonia_coenia_GV             TATGATAGCC TTGCGGAGAA AGATCAGAAT GAACAACGAA ---------- -----GCCGT CATCGACACC CTGCAGACGC
193_Lymantria_obfuscata_NPV      AATGATTTGC ATCCGGCGAG CGGTGAAGAT GTGCGAC--- ---------- -----GCGCG CATCGCGTGC CTGCTCACGC
195_Trabala_vishnou_NPV          CATGGTGTGC ATCAAGCGCA CGGTGCGCCT GATCAAAACA AAA------- -----GTCAG AATTGCCGCT CTGCTCACGC
202_Lasiocampa_quercus_NPV       CATGATATGT CTTAAACGAA CGGCGGTCGT CAAGAGA--- ---------- -----GTCAA AATTAACGTG TTGTTAACTC
205_Kotochalia_NPV               AATGATTGGT TTAAAGAAAA GAATCATCTT GAAGTCCGGA ---------- -----GTAAA GATAAACGCG CTTTTAACAC
228_Hyphantria_cunea_GV          TATGATTATG TTGCGCAAAG TTGTGGAAAT TGATAAAGAA ---------- -----ATTAA AATTAACACA CTTCAAACTC
248_Euplasia_lucipara_GV         AATGGTTATG TTGAAGAAGG TGGTGTGTGT GAACCAAACG ---------- -----GTAAA CATTAACACT TTGCAGACCG
254_Natada_nararia_GV            TATGGTTATG TTAAAAAAAA CAGTAGTTAT TGATAACACA ---------- -----ATTAA AATTAATACT TTACAAACGG
26_Melanchra_persicae_GV         GATGGTAATG TTGAAGAAAA TAGTTTGTAT TGATGACAAA ---------- -----ATTAC GATCAACACG CTACAAACGG
274_Mamestra_brassicae_NPV       AATGATTTGT ATAAAGAAAA AACTTTTAAT ACACGACATG CAATCGAACA CGACGGCAAC AATAGCTGCC CTACTAACTC
28_Samia_ricini_NPV              AATGATATGC ATCAAAAAAA AGTTGCGGGT GAACAACGAA ---------- -----GCCGA CATTAACGTA TTGTTGACAC
291_O._brumata_NPV_dartmoor      AATGATTTGT TTGAAGAAAA CAGTACAAGT GGATCAACAT ---------- -----GTAAA AATTATTACG CTATTGACGC
30_Antherea_pernyi_NPV           AATGATATGC ATCAAAAAAA AGTTGCGGGT GAACAACGAG ---------- -----GCCGA CATTAACGTA TTGTTGACAC
302lef8c                         CATGGTGTGT ATTAAAAAAT GcTTGCGTGT CACCAAGGAA ---------- -----GCGGA AATTAATGCG CTTTTGACGC
32_Kotochalia_NPV                AATGATTGGT TTAAAGAAAA GAATCATCTT GAAGTCCGGA ---------- -----GTAAA GATAAACGCG CTTTTAACAC
342_Wiseana_GV                   AATGGTCATG CTTCGCAAAG TGGTGTGCAT AGTCGATGGG ---------- -----GTTGA CCTAAACACC CTACAAACCG
344_wiseana_cervinata_NPV        CATGCTTTTA GTAAAGAAGC GCGTGTCCCT AAAGGACGGC ---------- -----GTAGA GATCAACTGC AACCTCACGC
353_Nymphalis_io_NPV             AATGGTTTGT TTACGACGAA CAGTCGTCAT TTCACATAAT GAC------- -----GTCAA AATTAACGTT CTCTTGACGC
362_Shinx_ligustri_NPV           AATGATATGC ATCAAAAAAA AGTTGCGGGT GAACAACGAA ---------- -----GCCGA CATTAACGTA TTGTTGACAC
367_Vanessa_cardui_NPV           AATGGTTTGT TTACGACGAA CAGTCGTCAT TTCACATAAT GAC------- -----GTCAA AATTAACGTT CTCTTGACGC
37_cacoecia_murinana_GV          TATGGTAATG TTAAAAAAAA GTGTTTGTAT CAATAACGAA ---------- -----ACACA CATTAACACT TTGCAACCTG
'380 Polygonia c-album NPV'      AATGGTTTGT TTACGACGAA CAGTCGTCAT TTCACATAAT GAC------- -----GTCAA AATTAACGTT CTCTTGACGC
384_Pieris_brassicae_GV          AATGGTGATG TTGAAAAAAG TGGTGTGTAT AGACAACGAG ---------- -----ATTCG TATAAACACT TTACAAACAC
402_idea_virgularia_NPV          TATGGTGTGC ATAGCTAAAA AGGTGCGCAT ATACGCGCCC GTTAAC---- -----GCTAC GATTAATGCG CTACTAACGC
436_Pachypasa_papyri_NPV         TATGATTTGT CTAAAACGAA CGGTAAAAAT TTGCGAAAAC ---------- -----ACCAC AATCAACGTG CTTTTGACTC
440lef8c                         CATGATATGT ATACGCAAGC AGGTTCGAAT AAATGAAAAA ---------- -----GTAAA CATTTCGGCG CTGCTCACTC
460_Bombyx_mori_NPV              AATGGTATGC ATCAAAAAAA AGTTGCGGGT GAACAACGAA ---------- -----GCCGA CATTAACGTG TTGTTGACAC
47_Actias_selene_NPV             TATGGTGTGC GTGGCCAAAA AATTACGCAT CCTGGCGCCG GTGCAC---- -----GCCAC GGTGCATGCG CTGCTCACCC
505_buzura_edwardsiiNPV          TATGATCTGT ATAAAGAAAA GATTAAAGAT TAACAGCGAT GTGGGAATC- -----GTAAA AATAATGACA TTGTTGACGC
54_Vanessa_atalanta_NPV          AATGGTTTGT TTACGACGAA CAGTCGTCAT TTCACATAAT GAC------- -----GTCAA AATTAACGTT CTCTTGACGC
545_Darna_trima_GV               TATGATTATG CTTAAAAAGG TAGTACGCAT CAACAACGAA ---------- -----ATACG CATTAATACA TTACAAACTC
595lef8c                         AATCCTTTGT TTGAACAGGC TAGTGCTTTT TTTCAAAGAC CCCCTCGAT- -----ATTTC AGTGTATACT TTATCGACGC
63_Aroa_discalis_NPV             CATGGTATGC ATTAAGCGCA CCGTGCTCAT GAAAAACACA AAC------- -----GTCAA GATTGCCGCG CTTTTGACGC
64_Achaea_faber_NPV              CATGATTTGT ATCAAAAAAC GATTGAAAAT CGCACAAGAC AAGGATAAAG ATCCCGTTCA CATTGCCACG TTGCTAACTC
66_sp_littoralis_GV              TATGGTGATG TTAAAGCGAA CTGTGCTTAT CGATGATGAT ---------- -----GTTAT TATTAACACT TTGCGACCGG
67_Orgya_mixta_NPV               CATGGTATGC ATTAAGCGCA CCGTGCTCAT GAAAAACACA AAC------- -----GTCAA GATTGCCGCG CTTTTGACGC
692lef8c                         AATGATTTGT TTGAAGAAAC AAATCAAAAT TAAAAACAAC GACAAAAAC- -----ATTTC AATAAATACT TTATTAACGC
744lef8c                         TATGGTGATG TTAAAAAAAA GTGTTCGTAT CCATAACGAA ---------- -----ATACG CCTAAACACC TTGCGTACTG
750lef8c                         AATGATTTGC ATCAAGCGTG TGGTGAAGCT GCGCGGC--- ---------- -----GTGCA CATTGCGTGC CTGCTCACGC
766_Sp_litura_GV                 TATGGTGATG TTAAAGCGAA CTGTGCTTAT CGATGATGAT ---------- -----GTTAT TATTAACACT TTGCAAACGG
768_Euproctis_similis_NPV        CATGGTGTGC ATCAAAAAGC AAATGAACAT AGACAGCGCT GGT------- -----GTGAA AATCAACGTG TTGCTCACGC
78_Catopsila_pomona_NPV          AATGGTGTGC CTCAAAAAAC GTTTACACAT GACAGACAAC AAC------- -----GTGAA AATAAACGTG TTGCTTACAC
789_Heliconius_erato_NPV         CATGGTGTGT ATAGCCAAAA GAGTGCGCAT AGGGGCGCCC GTGAAC---- -----GCGGC AATCAACGCG CTGCTGACAC
798_Diacrisia_obliqua_NPV        CATGGTGTGT ATAGCGAAGA AGGTGCGCAT ATTAGCACCG GTGGAC---- -----GCCGC AATCAACGCG CTGTTGACGC
A1_Mythmnia_unipunctata_NPV      CATGATTTGC ATCAAGAAGC GACTACGCAT CACGAACGGC GACGAAAGC- -----GCGAC GATATCCGCG CTGCTCACGC
a10_Euproctis_chrysorrohea_NPV   CATGGTGTGC ATCAAAAAGC AAATGAACAT AGACAGCGCT GGT------- -----GTGAA AATCAACGTG CTTCTCACGC
a2_Mythmnia_separata_NPV_Japan   AATGATTTGT ATAAAGAAAA AACTTTTAAT ACACGACACG CAACGGAACA CGACGGCAAC AATAGCTGCC CTACTAACTC
a3_Mythmnia_separata_NPV_china   CATGATCTGT ATCAAGAAGA CGATTCGTAT GGAGCGCGAA CACGAGCACG AC---ACGCG CATAGTCGCT CTGCTCACGC
a4_Mythmnia_separata_NPV_India   CATGATCTGT ATCAAGAAGA CGATTCGTAT CGAGCGCGAA CACGAGACCG AC---ACGCG TATAGTCGCT CTGCTCACGC
a5_Cerapteryx_graminis_NPV       CATGATTTGT ATCAAAAAGA AGCTACGAAT TCGAGCGGAC GCGTCGAGCG GG---GCCAG CATTGTCGTT TTGCTGACAC
a6_Lacanobia_oleracea_GV         GATGGTAATG TTGAAGAAAA TAGTTTGTAT AAACAACAAA ---------- -----ATCAC TATCAACACT CTACAAACAG
a7_Aglais_urticae_NPV            AATGGTTTGT TTACGACGAA CAGTCGTCAT TTCACATAAT GAC------- -----GTCAA AATTAACGTT CTCTTGACGC
a9_Orgya_antiqua_NPV             TATGGTGTGC ATCAAAAGAA CAGTTGTCAT CAACAACACA AAC------- -----GTAAA AATTGTTGTG CTGCTGACAC
AcMNPV                           AATGATATGC ATCAAAAAAA AGTTGCGGGT GAACAACGAA ---------- -----GCCGA CATTAACGTA TTGTTGACAC
B10_Anticarsia_gemnitalis        TATGGTATGT GTGGCGAAAC GGCTGCGCAT ACCTTCC ACC CCCAAC---- -----GCTAC GGTCAACGCG TTGCTTACGC
b3_Plodia_interpunctella_GV      AATGGTTATG AAAAAGAAGA TAGTGTGCAT CACCAACAAC ---------- -----GTTCA TATTAACACT CTCCAAACAA
b4_Lymantria_monacha_NPV         CATGATTTGC ATCAAACGGC GGGTGAAGCT GCGCGCC--- ---------- -----GCCAC GATCGTGTGT TTGCTCACGC
b6_Orgya_leusostigma_NPV         AATGGTGTGC ATTAAACGGA CAGTGTTAAT GAAAAACCCA AAC------- -----GTCAA AATCGTCGTA TTGTTGACAC
BmNPV                            AATGGTATGC ATCAAAAAAA AGTTGCGGGT GAACAACAAA ---------- -----GCCGA CATTAACGTG TTGTTGACAC
c2_Trichoplusia_ni_NPV           AATGATATGC ATCAAAAAAA AGTTGCGGGT GAACAACGAA ---------- -----GCCGA CATTAACGTA TTGTTGACAC
c3_Potato_tuber_GV_8             CATGATCATG CTGAAACGAG TTGTGTGTAT CAATAACAA ---------- -----GTGCG AATAAACACG TTGCAAACAC
c4__Potato_tuber_GV_9            CATGATCATG CTGAAACGAG TTGTGTGTAT CAATAACAA ---------- -----GTGCG AATAAACACG TTGCAAACAC
C5_Ephasia_cautella_NPV          TATGGTATGT TTAAAAAAAT GTTTACTAAT TAATAACAT ---------- -----GTCAG GATTAATGCA CTCCTTACGC
CpGV                             TATGGTGATG TTGAAGCGGG TGGTGTACAT CAATAACA A ---------- -----ATCCA CATAAAGACG TTGCAGACCG
242_pieris_napi_GV               AATGGTGATG TTGAAAAAAG TGGTGTGTAT AAATAACA A ---------- -----ATTCG TATAAACACT TTACAAACAC
95_pieris_rapae_GV               AATGGTGATG TTGAAAAAAG TGGTGTGTAT AAATAACA A ---------- -----ATTCG TATAAACACT TTACAAACGC
```

```
CuniNPVlef8                          CATGCTGACC AAATGTATGG CCAACACGGG TCATATCGAC ---------- -----TACAC GGCGATGGAA ATTTATTCAC
EppoNPV                              CATGGTGTGT GTGGCTAAAA AATTGCGCAT ATTTCGACCT GTTCAC---- -----GCCAC AATCAACGCG CTGCTCACCC
HaSNPV                               TATGGTGTGC ATCAAACGGA AACTTAAAAT ACCCCAT--- ---------- -----GTGCA CATTACAGCG CTGCTCACCG
HzSNPV                               TATGGTGTGC ATCAAACGGA AACTTAAAAT ACCCCAC--- ---------- -----GTGCA CATTACAGCG CTGCTCACCG
i3_Panolis_flammea_NPV               AATGATCTGT ATAAAGAAAA AACTTTTGAT CCACGATATG CAATCGAATA CCCCGCCAC AATAGCTGCC CTGCTCACTC
K10_Sp_frugiperda_NPV                TATGATCTGT ATAAAGAAAA GATTAAAGAT TAACAGCCAT GTCCCAATC- -----GTAAA AATAATGACA TTGTTGACGC
K11Spodoptera_exempta_NPV            CATGATTTGT ATCAAAAAAC GATTGAAAAT CCTACAACAC AAGCATAAAG ATCCCGTTCA CATTGCCACG TTGCTGACTC
K12_Malacosoma_neustria_NPV          CATGATTTGT TTGAAACGTA CGGCGCTAGT GAAACAG--- ---------- -----ACCAA GATCAACGTG CTGCTCACCC
K1lef8c                              AATGATATGC ATCAAAAAAA AGTTGCGGGT GAACAACCAA ---------- -----GCCGA CATTAACGTA TTGTTGACAC
K2lef8c                              AATGATATGC ATTAAAAAAA AGTTGCGGGT GAACAACCAG ---------- -----GCAAA CATTAACGTA TTATTGACAC
K3lef8c                              CATGATTTGT TTGAAACGTC GGGCGCTAGT GAAACAG--- ---------- -----ACCAA GATCAACGTG CTGCTCACCC
K8_sp_sunia_NPV                      CATGGTGTGC ATCAAACGAG TGATCAATAT CGCGGGC--- ---------- -----AAAAA AATCACTACT TTATTGACGC
K9lef8c                              AATGATTTGT ATAAAGAAAA AACTTTTAAT ACACGACATG CAATCGAACA CGACGGCAAC AATAGCTGCC CTACTAACTC
LdMNPV                               AATGATTTGC ATCAAGCGCG CGGTGAAGCT GCCCGAC--- ---------- -----GTGAG CATCGCGTGC CTGCTCACGC
OpMNPV                               CATGGTGTGC ATAGCCAAAA GGGTGCGCAT TCGCCCCGCT GTCAAC---- -----GCTTC CATTAACGCG TTGCTGACGC
PxGV                                 TATGATTACT CTGAAAAAAA CCGTTAAAAT TTCCGACAAA ---------- -----ATAAT GATTAACACG TTACAGACGC
SeMNPV                               CATGATATGT ATCAAGAAAA AAGTTAAATA TCCCCACCCA GCCCCCCTC- -----GTCAA GATCATGGCG CTGCTTACGC
Sp_littoralis_NPV                    CATGATATGT ATACGCAAGC AGGTTCGAAT AAATGAAAAA ---------- -----GTAAA CATTTCGGCG CTGCTCACTC
Splitura_NPVgenome                   CATGATATGT ATACGCAAAC ATATTCTAAT AAACGACAAA ---------- -----GTTAA CATTTCGGCT CTGCTCACTC
XcGV                                 TATGGTAATG TTAAAACGAG TGGTTCGCGT ATTTAAAAAC ---------- -----GTTTA CATCAACACC CTTCAAACTC
M2lef8c                              AATGATAATG TTAAAAAAAG TCGTATGTAT CAACAACGAC ---------- -----ATTCG TATTAAAACT TTACAAACAC
N1lef8c                              CATGGTGTGT ATAGCAAAGA AGGTGCTCTT NTTACCACCG GGGGAC---- -----GCCGC AATCAACGCG CTGTTGACGC
'P1lef8c=E1110'                      CATGGTGATG CTCAAACAGC CCGTACAAAT AAACAACACA ---------- -----CACAC GGGGCACACT TTACAAACCG
CfMNPVlef8.txt                       CATGGTGTGC ATAGGTAAAA AGGTGCGCAT TTCCGCGCCC GTAAAC---- -----GCCAC CATTAACGCG CTCCTGACCC
Nele_NPV                             GTTGATAATG AAAAAAGCAA GATTGTTCGT AGACGAC--- ---------- -----GGATA CATCGATACA TATGTAACAC

10_Lasiocampa_quercus_NPV            CCTACGAATA CCACTATAAA CATTCTATAT ACTACAATCC CATT------ ---------- -----ATTCA GTGTACGATC
106_aporia_crataegi_NPV              CGTTTGAATA CCATTATAAA CGTTCGATTT ATCACAACAA ATTT------ ---------- -----AGCAA TTGCAGAATA
112_Abraxas_grossulariata_NPV        CGTACGAGTA CCACTACAAA AACTCGCTTT TCCACGCGCC CAATGTC--- ---------- -----GGCGC GTGCGAAATC
126_Spilosoma_lutea_NPV              CGTTCGAGTA TCACAACACC AAAAGTGTGG TAAATTTGCT GTCGCCG--- ---------- -----GTGCT GCAATTTTGC
14_Anomis_sabulifera_NPV             CCTACGAATA TCATAATGCT CAGTCTATTT TAAAAACAAA GCCC------ ---------- -----CATAT TGCATCCGTG
174_Erranis_defolaria_NPV            CGTTCGAGTA TCACAATAAT GATTCTATTC TGCGCAGCAC CACA------ ---------- -----GATGT GGGCGTCATT
175_Anomala_insulata_NPV             CCTACGAATA CCATTACAAG GATTCGCTGT TTCACAACCC GGCGATC--- ---------- -----GGTGG GTGCGAAATT
19_Junonia_coenia_GV                 CTTACGAATA CCACAACGAT AACGGCGTTT TAAAGCGCAA A--------- ---------- -----ATGTT CGAAATCGTC
193_Lymantria_obfuscata_NPV          CGCACGAGTA CCACAATTCG TCGGCGCTGC TCCACGCGGC GCC------- ---------- -----CTACG TGGGCCGTGT
195_Trabala_vishnou_NPV              CGTACGAATA CCACAACGAG AGCGCCTTGC TGAAGATCAC GAAC------ ---------- -----GACGT TGGCGTGATA
202_Lasiocampa_quercus_NPV           CGTTCGAGTA TCACAACGAC ATGTCTATAT TGAACCGCAA AGATGCACAA ---------- -----TTAAG CGTGGAGTTG
205_Kotochalia_NPV                   CCCACGAGTA CCATAACGAG TGTTCGCTAC TCAAAACGGC A--------- ---------- -----AATTT GGGCCATATC
228_Hyphantria_cunea_GV              CATTTGAGTA TCATAATGAA AAAAGTGTGG TGAACATGCT GTCA------ ---------- -----GTTTT GGAATTTTGC
248_Euplasia_lucipara_GV             CGTACGAGTA CCACAATACG AATAGTGTCG TGAACAAGCT T--------- ---------- -----GATTC GTTACATTTT
254_Natada_nararia_GV                CGTATGAATA CCACAATTCT TTAAGTTTTA TACACAATTA TAACAAAGTG ---------- -----CGTTT TGAAAACAGA
26_Melanchra_persicae_GV             CATACGAGTA TCATAGCGCA AATAGTATAA TCACCAAACT T--------- ---------- -----GGTGA GTTAGAGTTT
274_Mamestra_brassicae_NPV           CGTACGAATA TCATAATAGC GATTCGATAC TGAACAAAGT A--------- ---------- -----GCCGG TTTGCAGATT
28_Samia_ricini_NPV                  CCTACGAATA CCACTATAAA CATTCTATAT ACTACAATCC CATT------ ---------- -----ATTCA GTGTACGATC
291_O._brumata_NPV_dartmoor          CGTTCGAGTA TCATCATGAC AATAGCGTGT TGAGACAGAT C--------- ---------- -----AAAAA TGTTGAAATT
30_Antherea_pernyi_NPV               CCTACGAATA CCACTATAAA CATTCTATAT ACTACAATCC CATT------ ---------- -----ATTCA GTGTACGATC
302lef8c                             CTTAcGAGTA CCATTACAAC CATTCCGTGT ACCACAACCC TGAT------ ---------- -----ATAAA ATGTACAATT
32_Kotochalia_NPV                    CCCACGAGTA CCATAACGAG TGTTCGCTAC TCAAAACGGC A--------- ---------- -----AATTT GGGCCATATC
342_Wiseana_GV                       CGTACGAGTA TCACAATAAA AATAGTATAG TAAACAGCAA T--------- ---------- -----GCGGC GCTAAGGTTT
344_wiseana_cervinata_NPV            CCTACGAGTA TCACCACGCC AACAGCGTTT TTCGGAGCGG AGGCGGCGGC GGCGAG---- -----GGATG GCCGCGGATA
353_Nymphalis_io_NPV                 CGTACGAATA TCACAATTCG ATGTCTTTAT TAAACACACA TGGCGAC--- ---------- -----CAAAT CGGTCACATT
362_Shinx_ligustri_NPV               CCTACGAATA CCACTATAAA CATTCTATAT ACTACAACCC CATT------ ---------- -----ATTCA GTGTACGATC
367_Vanessa_cardui_NPV               CGTACGAATA TCACAATTCG ATGTCTTTAT TAAACACACA TGGCGAC--- ---------- -----CAAAT CGGTCACATT
37_cacoecia_murinana_GV              CTTACGAATA CCACAACAAA CATAGTGTAA TTAACAACGA CTCTGTTTTA ---------- -----AATTT TGAGGATAAA
'380 Polygonia c-album NPV'          CGTACGAATA TCACAATTCG ATGTCTTTAT TAAACACACA TGGCGAC--- ---------- -----CAAAT CGGTCACATT
384_Pieris_brassicae_GV              CGTACGAGTA TCACAACGAA AACAGTTTGG TAAACAGAGA TCAAGTG--- ---------- -----TTACA ATTTGAAAAT
402_idea_virgularia_NPV              CGTATGAGTA TCATTACAAA AATTCGCTGT TTCACACGCC GATTATT--- ---------- -----GGCGC GTGTAAAATT
436_Pachypasa_papyri_NPV             CGTTTGAGTA TCACAATGAA ATGTCCATAA TACGAGATAA AATTACACCG TCATCACTAA CACAGCATAT AGTGGAACTT
440lef8c                             CGTACGAGTA CCACAACGTC GAATCTATCA TTGGCCAG-- ---------- ---------- -----ATCGG AGCCCGGATC
460_Bombyx_mori_NPV                  CCTACGAATA CCACTATAAA CATTCTATAT ACTACAATTC CATT------ ---------- -----ATTCA ATGTACGATC
47_Actias_selene_NPV                 CGTACGAGTA CCATTACAAG AGTTCGCTGT TTCACGACTA NGGAGTG--- ---------- -----GGCAA ATACGAAATC
505_buzura_edwardsiiNPV              CGTACGAATA CCACAATGAA TTGTCGTACT TGCGAACCGT G--------- ---------- -----CCCGG CATATGCGTC
54_Vanessa_atalanta_NPV              CGTACGAATA TCACAATTCG ATGTCTTTAT TAAACACACA TGGCGAC--- ---------- -----CAAAT CGGTCACATT
545_Darna_trima_GV                   CTTTCGAATA CCACAATTCT CAAAGTGTAA TTCACGCTCC CGAAATC--- ---------- --GATATTGC GTTCGAAGAT
595lef8c                             CATACGAATA TCATGGTGCG CGGTCAATAt tGAAAACtAT A--------- ---------- -----GAtAT TGGAGAAGTG
63_Aroa_discalis_NPV                 CCTACGAGTA CCACAACTCT GCGTCCATAT TGAACAGCTT GGAC------ ---------- -----GACGT GGGCACGGTC
64_Achaea_faber_NPV                  CGTACGAGTA CCACAATCAA TTGTCATATT TGCACACGTT C--------- ---------- -----GAAGG CGTAAGTGTC
66_sp_littoralis_GV                  CTTGCGAATA TCACAATTGT AATAGTTTTA TTACAAAACT CAAA------ ---------- -----GATTT AAAATTTTTG
67_Orgya_mixta_NPV                   CCTACGAGTA CCACAACTCT GCGTCCATAT TGAACAGCTT GGAC------ ---------- -----GACGT GGGCACGGTC
692lef8c                             CATACGAATA TCATAATTCT CGTTCGATAT TAAAAACTAC A--------- ---------- -----GATAT TGGAGAAGTG
744lef8c                             CTTTTGAATA CCTCATCAAA CATAGTGTCA TTAACAAAGG CAGTGTTT-- ---------- -----TTAGT TTTGAACATT
750lef8c                             CGCACGAGTA CCACAATTCG GCGTCGCTCG TCAACACGAC GCCA------ ---------- -----TACGT GGGCCGTGTT
766_Sp_litura_GV                     CTTGCGAATA TCACAATTGT AATAGTTTTA TTACAAAACT CAAA------ ---------- -----GATTT AAAATTTTTG
768_Euproctis_similis_NPV            CGTACGAGTA CCATTACAAA AAATCCATTT AC[illegible] ---------- -----TCTGG GTGTCGCGTG
78_Catopsila_pomona_NPV              CGTACGAATA CCATTACAAA GATTCCATTT AT[illegible] ---------- -----GCGGC TTGTCAAATT
789_Heliconius_erato_NPV             CGTACGAGTA CCACTACAAA AATTCGCTGT TT[illegible] ---------- -----AGCGC GTGCACCATC
798_Diacrisia_obliqua_NPV            CATACGAATA TCACTACAAA AACTCGCTGT TT[illegible] ---------- -----GGCGC GTGTCAAATA
A1_Mythmnia_unipunctata_NPV          CCTACGAGTA CCACAACTCG GACTCGATTC TG[illegible] ---------- -----GCGGC GCTCGTCGTC
a10_Euproctis_chrysorrohea_NPV       CGTACGAGTA CCATTACAAA AAATCCATTT AC[illegible] ---------- -----TCTGG GTGTCGCGTG
a2_Mythmnia_separata_NPV_Japan       CGTACGAATA TCATAATAGC GATTCGATAC TCAA[illegible] ---------- -----GCCGG TTTGCAGATT
a3_Mythmnia_separata_NPV_china       CCTACGAGTA CCACAATAGC GATTCTATAA TG[illegible] ---------- -----AACGG ACTGATCATC
a4_Mythmnia_separata_NPV_India       CCTACGAGTA CCACAATAGC GATTCTATAA TG[illegible] ---------- -----AACGG ACTGATCATC
a5_Cerapteryx_graminis_NPV           CGTACGAGTA CCACAACGAA CTGTCGTACC TG[illegible] ---------- -----CCCGG ACTCGAGATC
a6_Lacanobia_oleracea_GV             CATACGAGTA CCATAGCGCA AATAGTATTA TC[illegible] ---------- -----GGTCA GTTAGAGTTT
a7_Aglais_urticae_NPV                CGTACGAATA TCACAATTCG ATGTCTTTAT TA[illegible] --- ---------- -----CAAAT CGGTCACATT
a9_Orgya_antiqua_NPV                 CTTACGAATA CCATAACAGC GGTTCAATTT TG[illegible] ---------- -----GACGT GGGCAAAGTC
AcMNPV                               CCTACGAATA CCACTATAAA CATTCTATAT AC[illegible] ---------- -----ATTCA GTGTACGATC
B10_Anticarsia_gemnitalis            CCTACGAGTA TCATTACAAA AACTCGCTGT TT[illegible] ---------- -----AACGG GTGTCAAATT
b3_Plodia_interpunctella_GV          CTTTCGAGTA CCACAACGAA AGCAGCGTGG TC[illegible] ---------- -----AAATT CGTAGATAAA
```

```
b4_Lymantria_monacha_NPV                CGCACGAATA CCACAACTCG GCGTCGCTGC TCAACACGGT CAAC------ ---------- -----AATGT GGGGCGCGTC
b6_Orgya_leusostigma_NPV                CCTACGAGTA TCACAACGAC GCATCCATAT TGAACACGGT CGAC------ ---------- -----GACGT GGGTGTAATA
BmNPV                                   CCTACGAATA CCACTATAAA CATTCTATAT ACTACAATTC CATT------ ---------- -----ATTCA ATGTACGATC
c2_Trichoplusia_ni_NPV                  CCTACGAATA CCACTATAAA CATTCTATAT ACTATAATGC CATT------ ---------- -----ATTCA GTGTACGATC
c3_Potato_tuber_GV_8                    CGTTCGAATA TCACAACGAA AACAGTTTGG TGAACACGGA T--------- ---------- -----CCTGC GTTGATTTTT
c4__Potato_tuber_GV_9                   CGTTCGAATA TCACAACGAA AACAGTTTGG TGAACACGGA T--------- ---------- -----CCTGC GTTGATTTTT
C5_Ephasia_cautella_NPV                 CTTACGAATA CCATTATAAA GAGTCTATTT ATTTCAATGA TGCC------ ---------- -----AGCAA ATGTATAATA
CpGV                                    CGTACGAGTA CCACAACGAG CATAGTTTGT TGAATAGCGA GGATGTCCTG ---------- -----TTGAG GTGTGATGAG
242_pieris_napi_GV                      CGTACGAGTA TCACAACGAA AACAGTTTGG TAAATACAGA TCAAGTG--- ---------- -----TTACA ATTTGAAAAT
95_pieris_rapae_GV                      CGTACGAGTA TCACAACGAA AACAGTTTGG TAAATACAGA TCAAGTG--- ---------- -----TTACA ATTTGAAAAT
CuniNPVlef8                             GCTATTCACT CGTCAACCAA CTGATTGAAA ACTATGGTTA CGATATCGAG ---------- -----AAACC CCTCGATAGG
EppoNPV                                 CGTACGAGTA TCACTACAAA AGTTCGCTAT TTCATGACGC TGCTGTG--- ---------- -----AACGG ATGCCAAATT
HaSNPV                                  CATACGAGTA TCATAATTCC AATTCGATTT TGCCCACCGA CAAC------ ---------- -----GACGT GGCCACAATA
HzSNPV                                  CATACGAGTA TCATAATTCC AATTCGATTT TGCGCACCGA CAAC------ ---------- -----GACGT GGCCACAATA
i3_Panolis_flammea_NPV                  CGTACGAATA TCACAATAGC GATTCTATAC TAAACAAAGT G--------- ---------- -----GCCGG TTTGCAGATT
K10_Sp_frugiperda_NPV                   CGTACGAATA TCACAATGAA TTGTCGTACT TGCTAACCGT G--------- ---------- -----CCCGG CATATGCGTC
K11Spodoptera_exempta_NPV               CGTACGAGTA CCACAACCAA TTGTCATATT TGCACACGGT C--------- ---------- -----GAAGG CGTAAGTGTC
K12_Malacosoma_neustria_NPV             CGTTCGAGTA TCACAACTCT AGTTCTGTGA TCAACGACGG CTACGTCGAC ---------- -----GTCGA CGTAGTTCTG
K1lef8c                                 CCTACGAATA CCACTATAAA CATTCTATAT ACTATAACGC CATT------ ---------- -----ATTCA GTGTACGATC
K2lef8c                                 CCTCCGAATA CCCCTATAAA CATTCTATAT ACTATAATCC TATT------ ---------- -----ATTCA ATGTACCATC
K3lef8c                                 CGTTCGAGTA TCACAACTCT AGTTCTGTGA TCAAGGACGG CAACGTCGGC G--------- -----TCGAC CGTAGTTCTG
K8_sp_sunia_NPV                         CTTTCGAATA TCACAATGCT CATTCGATCC TCAAGAAATC TCCG------ ---------- -----GAAAT CGGTAAAATA
K9lef8c                                 CGTACGAATA TCATAATAGC GATTCGATAC TAAACAAAGT A--------- ---------- -----GCCGG TTTGCACATT
LdMNPV                                  CGCACGAGTA CCACAATTCG GCGTCGCTGC TCAACGCGGC GCCC------ ---------- -----GACGT GGGCCGCGTC
OpMNPV                                  CGTACGAGTA CCATTACAAA AACTCGCTGT TTCACGCGCC GACGGTC--- ---------- -----AGCGC GTGCGAAATC
PxGV                                    CCTTTGAATA TCACAACTCT GAAAGTATAG TCAGAAATTT CAAA------ ---------- -----ATGTT GAATTTTGTA
SeMNPV                                  CCTACGAGTA TCACAACGAG CTGTCGTATC TGCGCACGGT C--------- ---------- -----GACGG TATCGATGTC
Sp_littoralis_NPV                       CGTACGAGTA CCACAACGTC GAATCTATCA TTGGCCAG-- ---------- ---------- -----ATCGG AGCCCGGATC
Splitura_NPVgenome                      CGTACGAGTA CCACAACGGC GAATCTATTG TTGCTCAA-- ---------- ---------- -----ATCAA AGCTCGGATC
XcGV                                    CGTACGAATA TCACAATTTA AACAGTATGG TGCACACAAT GGAC------ ---------- -----GATCT AACGTTTGAA
M2lef8c                                 CATACGAATA TCATAACAAA GACAGTTTGG TAAATAAAGA TAATGTTTTT ---------- -----GTTTT TGAAGAAAAG
N1lef8c                                 CATACCAGTA TCACTACAAA AACTCGCTGT TTCACACGGC GACCGTC--- ---------- -----GGCGC GTGTCAAATA
'P1lef8c=E1110'                         CCTACGAGTA CCACAGGGCA AACAGGGTGG TCAACAAGAG T--------- ---------- -----AAAGT TTTAAAGTTT
CfMNPVlef8.txt                          CCTACGAATA TCACTACAAA AAGTCACTTT TCGACGCGGC GACAATC--- ---------- -----GGCGC GTGCAACATT
Nele_NPV                                CGTTCGAATT ACATCACGAA TCGTCGTGTT TC-------- ---------- ---------- -----AAAGA TAAAATTTTT

10_Lasiocampa_quercus_NPV               GTAGAA---A ACGATGACGT CAAATCGCTC ATGTCCAAGT TGGAGCAGTA TTATTATCGT AATTTTATAC ATTTATTTCA
106_aporia_crataegi_NPV                 ATCGAA---A ACGATAATGT AAAGTCTTTG ATGTCCAAAT TGGAACAGTA TCATTACCGC GATTATATCC ATTTGTTTTA
112_Abraxas_grossulariata_NPV           AAAGAA---A GCGACGACGT GCGCTGCCTC ATGTCCAAGC TGGAGCAGTA CTACTTTGCA AACTACACGC ACTTGTTTTA
126_Spilosoma_lutea_NPV                 GACCAG---C ACGACACCGT GGTTTCGCTC ACTTCTCACA TTATCCAAAC GTACTATAAA GATTTTATGC ACATATTTCA
14_Anomis_sabulifera_NPV                ATTGAA---A CAAACGACGT AACGGCGCTA ATGTCCAAAG TGGTGCAATA TTATTATAAA AATTACACCG ATATTTTTAA
174_Erranis_defolaria_NPV               CAGGAA---A ACGACGATGT GACATGCCTT ATGTCTAAAC TGGTACAATA CTATTACAAA AGCTATACAA ATATATTTTG
175_Anomala_insulata_NPV                GACGAA---A GCGACGACGT GCGCTGCCTC ATGTCCAAGC TAGAGCAGTA CTACTTTGCC GAGTACACGC ACCTGTTTTA
19_Junonia_coenia_GV                    GACGAC---G ACAACGACGT GACGTCGCTG CTCTCGCTGA TGATTCACCG TTACTATAGC AATTTTTTGC ATATTTTTCA
193_Lymantria_obfuscata_NPV             TTAGGAG--A CAGACGACGT GACCAGTCTC ATGTCCAAGT TGGTTCAATA CTACTACAAG AACCACACGA GCATNTTCAG
195_Trabala_vishnou_NPV                 GAGGAA---A ACGACGACGT CACGTGTCTA ATGTCGAAAT TGGTGCAATA CTATTACAAC GGCTACACCA ACATATTCAG
202_Lasiocampa_quercus_NPV              GAAGAG---A GCAATGATGT TCGTTGCCTC ATGTCCCAGT TGGTTGAGTA CTATTATAAA GACTACTCAC GCATATTTAG
205_Kotochalia_NPV                      GAGGAA---G AGGAGTATTC CAGTTGTCTC ATGTCGAAGT TGGTACAATA TTACTACAAA GACTATTTAC ATATCTTTCG
228_Hyphantria_cunea_GV                 GACAAC---G ACGACGACAT TGTTTCACTT ACGTCCACCA TTGTTCAAAA TTACTACAAA AATTATTTGC ACATTTTTTA
248_Euplasia_lucipara_GV                CGAGAC---G CGGACACCAC TGTCGCGCTT ATGTCCACAA TGGTGCGTCA ATACTACAAA GATTATCTAC ACATCTTTAA
254_Natada_nararia_GV                   AAAGAT---G AAAATACTAC AAATTGTTTA ATGTCTACTA TGATTTTAGA ATATTATAAA AACCACTTAC ACATTTTTAA
26_Melanchra_persicae_GV                AAAGAT---G AGGACACCAC AGTTTCGTTG ATGTCAACAA TGGTGAGACA ATATTATAAA GATTATCTTC ATATATTTAA
274_Mamestra_brassicae_NPV              GTTCAA---G ATCGTCACGT GTCTCAACTC ATGTCCAAAC TGCTTCAATA CTACTACGTG GGTTACATGG ACATGTTTGC
28_Samia_ricini_NPV                     GTAGAA---A ATGATGACGT CAAATCGCTC ATGTCCAAGT TGGAGCAGTA TTATTATCGT AATTTTATAC ATTTATTTCA
291_O._brumata_NPV_dartmoor             ATCGAA---A CCGACGATGT GTCGTGTCTC ATGTCCAAGA TGGTACAATA TTATTACAAG TCATATTTAC ATATATTCGC
30_Antherea_pernyi_NPV                  GTAGAA---A ACGATGACGT CAAATCGCTC ATGTCCAAGT TGGAGCAGTA TTATTATCGT AATTTTATAC ATTTATTTCA
302lef8c                                GTTGAA---A ATGATGACGT CAAATCATTA ATGTCAAAAT TGGAGCAATA CTATTACCGC AATTATATAC ATTTGTTTCA
32_Kotochalia_NPV                       GAGGAA---G AGGAGTATTC CAGTTGTCTC ATGTCGAAGT TGGTACAATA TTACTACAAA GACTATTTAC ATATCTTTCG
342_Wiseana_GV                          AAAGAG---C AAAACGATAC GACCTGCCTG ATGTCTACTA TGATTTTAAA ATATTACAAA GACTATTTAC ACATTTTTAA
344_wiseana_cervinata_NPV               ATCGAA---A AGGAAACGGT GGCGTCGCTC ATGTCGAAGC TCGTTCAGTA CTACTTCAAA GATTGCCTCC ACATTTTCGC
353_Nymphalis_io_NPV                    GAAGAA---A ACGACAATGT CACCTGTCTC ATGTCCAAAT TAGTTCAGTA CTATTACAAA GGTTTTACGA GAATTTTCAG
362_Shinx_ligustri_NPV                  GTAGAA---A ATGATGACGT CAAATCGCTC ATGTCCAAGT TGGAGCAGTA TTATTATCGT AATTTTATAC ATTTATTTCA
367_Vanessa_cardui_NPV                  GAAGAA---A ACGACAATGT CACCTGTCTC ATGTCCAAAT TAGTTCAGTA CTATTACAAA GGTTTTACGA GAATTTTCAG
37_cacoecia_murinana_GV                 TTGGGT---G AAAACACTAC AGCCGCTTTA ATGTCGACCA TGATTTTGGA ATATTACAAG AATTATTTGC ATATCTTTGG
'380 Polygonia c-album NPV'             GAAGAA---A ACGACAATGT CACCTGTCTC ATGTCCAAAT TAGTTCAGTA CTATTACAAA GGTTTTACGA GAATTTTCAG
384_Pieris_brassicae_GV                 AAAGCAAACG AAAACACTAC GGTTTCGTTA ATGTCCACCA TGATATTAGA GTATTACAAA AATTATTTGC ATATATTTGA
402_idea_virgularia_NPV                 GAGGAA---G AGGACGATGT GCGCTGTCTC ATGTCAAAAT TGGAGCAATA TTACTTTGCT AATTACACAC ATCTATTCTA
436_Pachypasa_papyri_NPV                GACGAAAAACA ATGACGATGT ACGATGTTTA ATGTCAAAAT TGGTACAATA CTATTACAAA GATTACACTC GCATATTCAG
440lef8c                                GTCGAC---G TCGATCATGT CTCTGCGCTA ATGTCCAAGA CGCTTCAGTA CTATTCCCGA TCGCATTTGC ACATTTTCGC
460_Bombyx_mori_NPV                     GTAGAA---A ACGATGACGT CAAATCGCTC ATGTCCAAAT TGGAGCAGTA TTATTACTGT AATTTTATAC ATTTATTTCA
47_Actias_selene_NPV                    GAGGAG---G GCGATGACGT GCGCTGCCTC ATGTCCAAGC TGGAACAGTA CTATTTTGCT GAGTACACGC ACCTGTTCTA
505_buzura_edwardsiiNPV                 GAAGAA---G AGAAACATTG TTCGCAATTA ATGTCCAAAC TGTTACACTA CTATTACAAA GATATACTAT CCATGTTTGC
54_Vanessa_atalanta_NPV                 GAAGAA---A ACGACAATGT CACCTGTCTC ATGTCCAAAT TAGTTCAGTA CTATTACAAA GGTTTTACGA GAATTTTCAG
545_Darna_trima_GV                      ATGCCGGGCG AAAATACAAC CGCGTCGCTC ATGTCTACTA TGATTTTAGA CTATTACAAA AACCATCTTA ACATATTCGA
595lef8c                                TCTaAA---A ATAAAAaCAt TGCATGTTCA AAGTCAGCAC TCGTTCAAGA TTATTAcAAA AAAGTAAATC ATATCTTTTC
63_Aroa_discalis_NPV                    GTGGAA---A ACGACGACGT CACCTGTCTC ATGTCCAAGC TGGTGCAATA CTATTACAAG AATTTCACAA ACATTTTTTA
64_Achaea_faber_NPV                     GAAGAA---A CTGTGGACGT TTCGCAGCTA ATGTCCAAAT TGTTGCAGTA CTACTACAAA GACCATCTTT CAATGTTCGC
66_sp_littoralis_GV                     CCCCAA---A ATGGATATGT GACTTCGCTT ACGTCCGCAA TGATACATCA TTACTATAAA GACTATTTCA ATATTTTTCC
67_Orgya_mixta_NPV                      GTGGAA---A ACGACGACGT CACCTGTCTC ATGTCCAAGC TGGTGCAATA CTATTACAAG AATTTCACAA ACATTTTTTA
692lef8c                                TCTAAA---A ATAAAAACAT AACATGTTTA ATGTCTGGAT TAGTTCAATA TTATTATAAA AATTTTATTC ATATCTTTTC
744lef8c                                AATGGT---G AAAACACTAC AGCGGCTTTA ATGTTTACCA TGATTTTGGA ATATTACAGG AATTATTTGC ATGGTTATGA
750lef8c                                gAGGAA---A ACGACGACGT GACCAGTGTC ATGTCCAAGC TGGTTCAGTA CTATTACAAA AACCACACAA GCATTTTTAG
766_Sp_litura_GV                        CCCCAA---A ATGAAGATGT GACTTCGCTT ACGTCCGCAA TGATACATCA TTACTATAAA GACTACTTCA ATATTTTTCC
768_Euproctis_similis_NPV               TTCGAA---A ACGACGACGT CACAGCGCTC ATGTCCAAAT TGGAACAGTA TTATTACAAT GATTTTGTGC ACCTGTTTCA
78_Catopsila_pomona_NPV                 GCAAAC---A AAGACGATGT GAAATCGTTA ATGTCCAAAT TGGAGCAATA TTATTACAAC AATTTTGTGC ATTTGTTTCA
789_Heliconius_erato_NPV                GAGGAG---A ACGACGACGT GCGTTGCCTT ATGTCCAAAC TGGAGCAGTA CTATTTTGCC GAGTACACGC ATTTGTTTTA
798_Diacrisia_obliqua_NPV               GTAGAA---A ACGACGACGT GCGCTGCCTC ATGTCAAAAT TGGAACAGTA TTATTTTGCC AATTACACGC ACTTGTTTTA
A1_Mythmnia_unipunctata_NPV             GAGGAG---A CGCCGCACGT TTCGCAGCTG CTATCGAAAC TGCTGCAGTA CTACTACCAC GATCACCTGT ACATGTTTGC
a10_Euproctis_chrysorrohea_NPV          CTCGAA---A ACGACGACGT CACAGCGCTC ATGTCCAAAT TGGAGCAGTA TTATTACAAT GATTTTGTGC ACCTGTTTCA
```

```
a2_Mythmnia_separata_NPV_Japan      GTTCAA---G ATCGTCACGT GTCTCAACTC ATGTCCAAAC TGCTTCAATA CTACTACGTG GGTTACATGG ACATGTTTGC
a3_Mythmnia_separata_NPV_china      GAGGAG---A GCGATCATGT GTCGAATCTA ATGTCGAAGC TGGTCCAGTA CTACTATCGC AACTTTCTGC GCATGTTCGC
a4_Mythmnia_separata_NPV_India      GAGGAG---A GCGATCATGT GTCGAATCTA ATGTCGAAGC TGGTCCAGTA CTACTATCGC AACTTTCTGC GTATGTTCGC
a5_Cerapteryx_graminis_NPV          GAGGAG---A CGCCACACGT TTCGCAGCTC ATGTCTAAAC TGCTGCAGTA CTACTATCGT GATCACCTCA ACATGTTTGC
a6_Lacanobia_oleracea_GV            AAAGAT---G AGGACACCAC AGTTTCGTTG ATGTCAACAA TGGTGAGCCA ATATTATAAA GATTATCTTC ATATTTTTAA
a7_Aglais_urticae_NPV               GAAGAA---A ACGACAATGT CACCTGTCTC ATGTCCAAAT TAGTTCAGTA CTATTACAAA GGTTTTACGA GAATTTTCAG
a9_Orgya_antiqua_NPV                GTCGAG---C GCGACGACGT GACTTGCCTA ATGTCCAAGC CGGTGCAGTA TTATTACAAA AATTTCACAA ACATATTTTA
AcMNPV                              GTAGAA---A ACGATGACGT CAAATCGCTC ATGTCCAAGT TGGAGCAGTA TTATTATCGT AATTTTATAC ATTTATTTCA
B10_Anticarsia_gemnitalis           GAAGAA---A ATGATGACGT GCTGTGTCTC ATGTCAAAAC TTGAACAGTA TTACTTTACC GACTACACGC ACTTATTTTA
b3_Plodia_interpunctella_GV         CCCGGG---G AAGACAACAC AACGACACTG ATGTCCACTA TGATTCTGAA ATATTACAAG AATCATCTTC ACATATTCAA
b4_Lymantria_monacha_NPV            GACGAG---A ACGACGACGT AACGAGTCTC ATGTCCAAGT TGGTCCAATA CTACTACAAG AACCACACGG GCATCTTCAG
b6_Orgya_leusostigma_NPV            TTGGAA---C GCGACGATGT TACATGTTTA ATGTCCAAGC TGGTGCAATA CTATTACAAG AATTTCACAA ACATTTTTTA
BmNPV                               GTAGAA---A ACGATGACGT CAAATCGCTC ATGTCCAAAT TGGAGCAGTA TTATTACTGT AATTTTATAC ATTTATTTCA
c2_Trichoplusia_ni_NPV              GTAGAA---A ACGATGACGT CAAATCGCTC ATGTCCAAGT TGGAGCAGTA TTATTATCGT AATTTTATAC ATTTATTTCA
c3_Potato_tuber_GV_8                GAAGAA---A ACAACACCAC CTGTCAACTC ATGTCGACCA TGATTCTGGA ATATTATCGC AACCATTTGC ACATTTTCAA
c4__Potato_tuber_GV_9               GAAGAA---A ACAACACCAC CTGTCAACTC ATGTCGACCA TGATTCTGGA ATATTATCGC AACCATTTGC ACATTTTCAA
C5_Ephasia_cautella_NPV             ATTGAA---A ATGATGATGT GACTTCGCTA ACTTCCAAAT TGGAACAATA TTATTTTAAA AATTTTATGC ATTTATTTTA
CpGV                                ACGGGACCCA TGAACACTAC GGCGTCGCTC ATGTCCACAA TGGTGTTGGA GTACTACAAG AACTATTTGC ACATATTCAA
242_pieris_napi_GV                  AAAGCAAACG AAAACACTAC GGTTTCGTTA ATGTCCACCA TGATATTAGA GTATTACAAA AATTATTTGC ATATATTTGA
95_pieris_rapae_GV                  AAAGCAAACG AAAACACTAC GGTTTCGTTA ATGTCCACCA TGATATTAGA GTATTACAAA AATTATTTGC ATATATTTGA
CuniNPVlef8                         ATGAAA---C GCGGCGGAGC TGAGGACTTT ATGTCCGTAT TCACACTGCG CACCCTCGGT CCAAGCATCG CGGAATTTCT
EppoNPV                             GAAGAA---A ATGATGACGT ATTGTGTTTA ATGTCAAAAT TGGAACAATA TTATTTTGCC GATTACACGC ATTTGTTTTA
HaSNPV                              ATCGAA---A AGGACACAGT CGTTTGTTTG ATGTCCAAAT TGGTGCAGTA CTATTATAAA GGATATACGC AAATGTTCTG
HzSNPV                              ATCGAA---A AGGACACAGT CGTTTGTTTG ATGTCCAAAT TGGTGCAGTA CTATTATAAA GGGTATACGC AAATGTTCTG
i3_Panolis_flammea_NPV              GTTCAA---G ATCGGCACGT GTCCCAACTC ATGTCCAAAC TGCTACAATA CTATTACGTG GGCTATATGG ACATGTTCGC
K10_Sp_frugiperda_NPV               GAAGAA---G AGAAACATTG TTCGCAATTA ATGTCCAAAC TGTTACACTA CTATTACAAA GATATACTAT CCATGTTTGC
K11Spodoptera_exempta_NPV           GAAGAA---A CTGTGGACGT GTCGCAGCTA ATGTCCAAAC TGTTGCAGTA CTACTACAAA GACCATCTTT CAATGTTCGC
K12_Malacosoma_neustria_NPV         CAGGAG---G ACGACGATGT GCGATGTCTA ATGTCCAAAT TGGTGCAATA TTATTACAAA AACTACACAA ACATATTCAG
K1lef8c                             GTAGAA---A ACGATGACGT CAAATCGCTC ATGTCCAAGT TGGAGCAGTA TTATTATCGT AATTTTATAC ATTTATTTCA
K2lef8c                             GTAGAA---A ACGATGACGT CAAATCGCTC ATGTCCAAAT TGGAACAGTA TTATTATCGC AATTTTATAC ATTTGTTTCA
K3lef8c                             CAGGAG---G AAGACGATGT GCGATGTCTA ATGTCCAAAT TGGTGCAATA TTATTACAAA AACTACACAA ACATATTCAG
K8_sp_sunia_NPV                     GAGGAG---A ACGATGATGT ATCTAGTCTC ATGTCGAAAC TGGTACAATA TTATTATTTT GGCTATACAC GAATCTTTAG
K9lef8c                             GTTCAA---G ATCGTCACGT ATCTCAACTC ATGTCCAAAC TGCTTCAATA cTACTACGTG GGTTACATGG ACATGTTTGC
LdMNPV                              GAGGAG---A GCGACGACGT GACCAGTCTC ATGTCCAAGT TGGTCCAATA CTACTACAAG AACCACACGA GCATCTTCAG
OpMNPV                              GAGGAG---A GCGACGACGT GCGCTGCCTC ATGTCCAAGC TGGAGCAGTA CTACTTTGCT GAGTACACGC ATTTGTTTTA
PxGV                                GACGAG---C ACGACGACAC CACCTCCTTG ATGTCGACCA TGATCAGAAC CTATTACAAT GGTTATTTAC AAATATTCAA
SeMNPV                              GAAGAG---G AGAGGCATTG CTCGCAGCTC ATGTCCAAGC TGCTACATTA TTACTACAAA GATATACTGT CCATGTTCGC
Sp_littoralis_NPV                   GTCGAC---G TCGATCATGT CTCTGCGCTA ATGTCCAAGA CGCTTCAGTA CTATTACCGA TCGCATTTGC ACATTTTCGC
Splitura_NPVgenome                  GTCGAC---G CTGATCATGT TTCTGCGCTA ATGTCCAAGA CGCTTCAATA CTATTACCGA TCGCATCTGC ACATTTTCGC
XcGV                                GACCTG---A AAGACGATGT GACGTGTCTA ATGTCAACCA TGATTCATTA CTATTATCGC AACTACTTGC ACATATTTCC
M2lef8c                             TCAGAC---G AAAACACTAC CACGTCGTTG ATGTCAACTT GTATTTTAGA ATATTACAAA AATTATCTTC ACATTTTTAA
N1lef8c                             GGGGAA---A ACGACGACGT GCGCTGCCTG ATGTCAAAAT TGGAACAGTA TTATTTTGCC AATTACACGC ACCTGTTTTA
'P1lef8c=E1110'                     ATACAA---C ACGACGACAC CACATCGCTC ATGTCCACAA TGATTTTACA ATACTACAGG GGTTTATTTG CAATATTCAA
CfMNPVlef8.txt                      GAAGAA---A ACGACGACGT TCGCTGCCTC ATGTCGAAGC TGGAACAATA TTACTTTGCC AAGTACACGC ATTTGTTCTA
Nele_NPV                            CTTAAC---G AATCTTGTGT GAGATCTCTG TTGTCTCGTA CTACGTTGTT TTATTGCAAT AAACATACGG ATGTTTTTAT

10_Lasiocampa_quercus_NPV           TACCACACCC GTACCCAAAT TAATTGTGTC CCTAACCAAT CTAAAGAATG CCATGCCTGT GTTTGAGTAC AAA
106_aporia_crataegi_NPV             TACAATTCCC GTGCCCAAAT TGATTGTGTC GTTGACCAAT TTAAAAAACG CTATGCCCGT TTTTAAATAT GAC
112_Abraxas_grossulariata_NPV       CACCATTCCT GTGCCCAAAA TGATAGTGGC CCTCACAAAT TTAAAGAACG CCATGCCTGT GCTGCGCTAC GGCTCG----
126_Spilosoma_lutea_NPV             CACCATTCCG TTGCCCAAAT TAATTGTGAG CCTGACCAAT CTAAAAAACG GCATGGTGGT GCAAAGTCCT TTG
14_Anomis_sabulifera_NPV            TACTGTGCCT GTGCCGAAAC TGATCGTGTC ACTGACCAAT TTGAAAAACG CCATGCCGGT GATTCGCTAC GAC
174_Erranis_defolaria_NPV           TACCGTACCA GTGCCTAAAT TAATAGTTTC GTTGACAAAT TTAAAAAATG CCATGCCCGT CATCAGTTAC AAC
175_Anomala_insulata_NPV            CACCATTCCA GTGCCCAAAA TGATCGTGGC CCTAACCAAC TTGAAGAATG CCATGCCCGT GCTGCCGTAC AGC----
19_Junonia_coenia_GV                CACGATACCC GTGGCGAAAT TGATCGTCAG TTTGACGAAC TTGAAAAACG GCATGGTGGT AAATCGCAGC AGC
193_Lymantria_obfuscata_NPV         CACTCTGCCC GTGCCCAAGC TGATCGTGTC GCTGACCAAC TTGAAGAACG CGATGCCCGT GCAGAGCTAC ACG
195_Trabala_vishnou_NPV             CACGGTGCCG GTGCCCAAGT TGATTGTGTC GTTGACCAAT CTGAAAAACG CCATGCCGGT GTTCGGCTAT TGC
202_Lasiocampa_quercus_NPV          CACGGTGCCC GTGCCCAAAC TAATTGTATC GTTGGCCAGC CTCAAAAACG CAATGCCCAT ATTGGCGTAC -AC
205_Kotochalia_NPV                  CACGTTGCCC GTGCCTAAGC TGATCGTGTC GCTGACCAAT TTGAAAAACG CCATGCCAGT GTTAAATTTC AAC
228_Hyphantria_cunea_GV             TTCGATTCCC TTACCCAAGT TAATTGTGAG CCTGACTAAT TTGAAAAACG GTATGGTTGT TCAAACACCT TTA
248_Euplasia_lucipara_GV            CACTATACCG GTCGCAAAAT TAATTGTTAG TTTAACGAAT CTAAAGAACG GTATGGTGGT GCAGACCGCG GAA
254_Natada_nararia_GV               CACCATTCCT GTTGCAAAAT TAATTGTTAG TTTAACAAAC TTAAAAAATG GCATGGTTGT TCAAAAAAAT AAC-------
26_Melanchra_persicae_GV            CACCATACCT GTGGCAAAAC TGATCGTTAG TTTGACTAAT TTAAAAAACG GTATGGTAGT ACAAACGCCC GAT
274_Mamestra_brassicae_NPV          AATAACGCCG GTTCCCAAAC TAATTGTTTC GGTGACGAAC CTCAAAAACG CAATGCCCGT CGTCGAATAT TCA
28_Samia_ricini_NPV                 TACCACGCCC GTACCCAAAT TGATTGTGTC CCTAACCAAT CTAAAGAATG CCATGCCCGT GTTTGAGTAC AAA
291_O._brumata_NPV_dartmoor         TACGACACCG ACACCAAAAT TAATTGTGTC GTTGACAAAC TTAAAAAATG CCATGCCCGT CATATCATAC GAC
30_Antherea_pernyi_NPV              CACCACGCCC GTACCCAAAT TGATTGTGTC CCTAACCAAT CTAAAGAATG CCATGCCCGT GTTTGAGTAC AAA
302lef8c                            TACCACGCCA GTGCCCAAAT TGATAGTGTC TTTGACAAAT TTAAAAAATG CTATACCTTG TTTAGAATAC AAG-------
32_Kotochalia_NPV                   CACGTTGCCC GTGCCTAAGC TGATCGTGTC GCTGACCAAT TTGAAAACCG CCATGCCAGT GTTAAATTTC AAC
342_Wiseana_GV                      CACCATACCC GTCTCAAAAC TGATCGTCAG TTTGACAAAT TTAAAAAATG GCATGGTCAT ACAAACGTCC GC--------
344_wiseana_cervinata_NPV           CCGAACGCCC GTGGCCAAGC TCATCGTCAC CTCCAACAAC CTCAAGAACG CCATGGTCGT GTTGAGGGCG CGC
353_Nymphalis_io_NPV                TACGATTCCA GTACCTAAAC TAATCGTTTC GTTGACAAAT CTAAAAAATG CCATGCCCGT TGTAGATTAT TGC
362_Shinx_ligustri_NPV              TACCACGCCC GTACCCAAAT TGATTGTGTC CCTAACCAAT CTAAAGAATG CCATGCCCGT GTTTGAGTAC AAA
367_Vanessa_cardui_NPV              TACGATTCCA GTACCTAAAC TAATCGTTTC GTTGACAAAT CTAAAAAATG CCATGCCCGT TGTAGATTAT TGC
37_cacoecia_murinana_GV             CTCTATTCCT GTGGCAAAAC TCATAGTTAG TTTGACCAAT TTAAAAAACG GCATGGTGGT CCTAAAAAAC AAT-------
'380 Polygonia c-album NPV'         TACGATTCCA GTACCTAAAC TAATCGTTTC GTTGACAAAT CTAAAAAATG CCATGCCCGT TGTAGATTAT TGC
384_Pieris_brassicae_GV             CACCATACCC GTGTCTAAAA TAATAGTTAG TTTGACAAAT TTAAAAAACG GTATGGTTGT GCAAAATAAA AAT--------
402_idea_virgularia_NPV             CACCATACCG GTACCCAAGA TGATAGTGGC GTTGACAAAT TTAAAAAACG CCATGCCGGT AATACGCTAC AGT----
436_Pachypasa_papyri_NPV            CACGGTTCCA GTGCCCAAAT TAATTGTGTC GTTGACTAAT CTTAAAAATG CTATGCCCGT GTTAGCCTAT AAT
440lef8c                            GACGATACCC GTGCCCAAAC TGATCGTGTC CGTGACAAAT TTGAAAAACG CAATGCCCGT GATTGAGTAC GAT
460_Bombyx_mori_NPV                 TACCACGCCC GTACCCAAAT TGATTGTGTC CCTAACCAAT CTGAAGAATG CCATGCCCGT GTTTGAGTAC AAA
47_Actias_selene_NPV                CACCATCCCC GTGCCTAAAA TGATCGTGGC GCTGACCAAC CTAAAAAACG CCATGCCCGT CATGCCGTAC AGT
505_buzura_edwardsiiNPV             TACGACGCCG GTGCCCAAGT CTATAGTTTC GTTGACCAAT TTGAAAAACG CCATGCCTGT CGTCGCGTAC AAT
54_Vanessa_atalanta_NPV             TACGATTCCA GTACCTAAAC TAATCGTTTC GTTGACAAAT CTAAAAAATG CCATGCCCGT TGTAGATTAT TGC
545_Darna_trima_GV                  TACCATACCC GTGGCCAAAC TTATTGTTAG TCTCACAAAC CTAAAAAATG GTATGGTAGT GCGCAAATCA AAC
595lef8c                            TACTATTCCC ACGCCCAAAt TGATAGTTTC CCTTCCCAAt tTAAaGAAAG CAATGCCTGT TGTtTCCTAT TAT
63_Aroa_discalis_NPV                CACGGTGCCG GTGCCCAAGT TGATTGTGTC GCTGACCAAC TTGAAAAACG CCATGCCCGT GCTGGTGCAC GAC
64_Achaea_faber_NPV                 CATCACGCCG GTGCCGAAAT TGATTGTTTC GCTGACAAAC TTGAAAAATG CCATGCCCGT CGTGGCCTAT GAT
66_sp_littoralis_GV                 CAGCATACCG GGAGCGAAGC TGATAGTGAG TTTGACAAAT CTAAAAAATG GTATGGTGGT ACAGTGTCCC CCT
67_Orgya_mixta_NPV                  CACGGTGCCG GTGCCCAAGT TGATTGTGTC GCTGACCAAC TTGAAAAACG CCATGCCCGT GCTGGTGCAC GAC
```

```
692lef8c                    AACTATACCC ATACCTAAAT TGATAGTTTC CCTAACCAAT TTAAAGAATG CAATGCCTGT TGTTTCTTTT TCAAAAAATA
744lef8c                    TTCTATACCT GTGGCAAATC TGATAGTTAG CTTGTCCAAT TTAAACAACG GCATGGTGGT CCTAAAAAAC AAA-------
750lef8c                    CACGTTGCCC GTGCCCAAGC TGATTGTGTG GGTGACCAAT TTGAAAAACG CCATGCCCGT CAAACGCTAN NNNNNNNNNN
766_Sp_litura_GV            CAGCATACCG GGAGCGAAGC TGATAGTGAG TTTGACAAAT CTAAAAAATG GTATGGTGGT ACAGTGTCCC ACTAAACTAT
768_Euproctis_similis_NPV   CACAATTCCC GTGCCCAAAT TGATCGTGTC TCTGACCAAC CTCAAAAACG CCATGCCCGT GTTCGCGTAC AACGAACAAA
78_Catopsila_pomona_NPV     CACGGTGCCG GTGCCAAAAT TGATTGTCTC TCTGACCAAT CTAAAAAACG CCATGCCCGT GTTTGAATAT AAAAAAAGCG
789_Heliconius_erato_NPV    CACCATTCCC GTGCCCAAAA TGATTGTGGC GCTGACCAAT CTGAAAAACG CCATGCCGGT GTTGCGTTAC GCGACGTCGC
798_Diacrisia_obliqua_NPV   CACAATTCCG GTACCAAAAA TGATAGTGGC TTTGACTAAT TTAAAAAACG CAATGCCCGT GTTGCGCTAC AGC-------
A1_Mythmnia_unipunctata_NPV GTCGACGCCG GTGCCCAAGT TGATTGTGTC GGTGACGAAT CTAAAAAACG CAATGCCCGT GCTGCGCTAC AATGGCATTC
a10_Euproctis_chrysorrohea_NPV CACAATTCCC GTGCCCAAAT TGATCGTGTC TCTGACCAAC CTCAAAAACG CCATGTCCGT GTTCGCGTAC AACGAACAAA
a2_Mythmnia_separata_NPV_Japan AATAACGCCG GTTCCCAAAC TAATTGTTTC GGTGACGAAC CTCAAAAACG CAATGCCCGT CGTCGAATAT TCAACTTTGC
a3_Mythmnia_separata_NPV_china CACGACGCCC GTGCCCAAAC TGATCGTGTC CGTCACGAAC CTGAAGAACG CCATGCCCGT GCTGCCGTAC AGGAGCGACT
a4_Mythmnia_separata_NPV_India CACGACGCCC GTGCCCAAAC TGATCGTGTC CGTCACGAAC CTGAAGAACG CCATGCCCGT GCTGCCGTAC AGGAGCGACT
a5_Cerapteryx_graminis_NPV  CACCACGCCG GTGCCGAAGC TCATAGTGTC GGTGACAAAT TTGAAAAACG CCATGCCCGT CGTGGCGTTC GACCCCCCCC
a6_Lacanobia_oleracea_GV    CACCATACCT GTGGCACAAC TGATCGTTAG TTTCNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNTTTTTTT
a7_Aglais_urticae_NPV       TACGATTCCA GTACCTAAAC TAATCGTTTC GTTGACAAAT CTAAAAAATG CCATGCCCGT TGTAGATTAT TGCGGTAGCA
a9_Orgya_antiqua_NPV        TACGGTGCCG GTGCCCAAGC TAATTGTGTC GCTGACCAAC TTGAAAAACG CCATGCCCGT AATGGCTTAT GACGAAAAGC
AcMNPV                      TACCACGCCC GTACCCAAAT TGATTGTGTC CCTAACCAAT CTAAAGAATG CCATGCCCGT GTTTGAGTAC AAA-------
B10_Anticarsia_gemnitalis   CACAATACCG GTGCCTAAAA TGATAGTAGC GCTGACCAAT TTAAAAAATG CAATGCCAGT GTTGAGTTAC AACAAGCTGG
b3_Plodia_interpunctella_GV CACTATCCCT GTTGCCAAAC TGATAGTCAG TCTGACCAAT CTCAAGAACG GAATGGTGGT TCAAAGAAAA GAC-------
b4_Lymantria_monacha_NPV    CACGCTGCCG GTGCCCAAGT TGATCGTGTC TCTGACCAAC TTGAAGAACG CCATGCCGGT GCAAAGCTAC ACGGACGCGC
b6_Orgya_leusostigma_NPV    CACGGTGCCG GTTCCCAAAC TGATCGTGTC GTTGACCAAT TTAAAAAATG CCATGCCTGT GCTCGTTTAC GACGACAACG
BmNPV                       TACCACGCCC GTACCCAAAT TGATTGTGTC CCTAACCAAT CTAAAGAATG CCATGCCCGT GTTTGAGTAC AAA-------
c2_Trichoplusia_ni_NPV      TACACAGCCC GTACCCAAAT TGATTGTGTC CCTAACCAAT CTAAAGAATG CCATGCCCGT GTTTGAGTAC AAACAGGAAA
c3_Potato_tuber_GV_8        CACGATTCCT GTGGCCAAAT TGATTGTCAG TCTGACCAAT CTAAAAAACG GCATGGTTGT ACGTCACACC GGACAAAGC-
c4__Potato_tuber_GV_9       CACGATTCCT GTGGCCAAAT TGATTGTCAG TCTGACCAAT CTAAAAAACG GCATGGTTGT ACGTCACACC GGACAAAGC-
C5_Ephasia_cautella_NPV     TACGCTACCG GTGCCCAAAC TTATAGTGGC GTTAACTAAT TTGAAAAATG CGATGCCTGT GTTCAGTTAC AGAAACAATG
CpGV                        CACGATTCCG GTGGCCAAGT TGATTGTGAG CCTGACCAAT TTGAAAAACG GCATGGTGGT GCAGAAGAAA GGT-------
242_pieris_napi_GV          CACCATACCC GTGTCTAAAA TAATAGTTAG TTTGACAAAT TTAAAAAACG GTATGGTTGT GCAAAATAAA AAT-------
95_pieris_rapae_GV          CACCATACCC GTGTCTAAAA TAATAGTTAG TTTGACAAAT TTAAAAAACG GTATGGTTGT GCAAAAAAAN AAT-------
CuniNPVlef8                 TTACATCCCA GCCTCGAAAT CCTTCGTCAC CATGTCCAAC CGCAAGAACC ACGTCCATGT GTTTCGGGTC GGTGAGTATG
EppoNPV                     CACCATTCCG GTGCCCAAAA TGATAGTGGC GCTGACTAAT TTAAAAAACG CCATGCCAGT TTTGAGTTAC AGCAAA----
HaSNPV                      CACCGTTCCG GTGCCTAAAC TGATCGTGTC ATTGACCAAT TTGAAAAACG CCATGCCGGT TTTAAGTTAT CAGGATTTGT
HzSNPV                      CACCGTCCCG GTGCCTAAAC TGATCGTGTC ATTGACCAAT TTGAAAAACG CCATGCCGGT TTTGAGTTAT CAGGATTTGT
i3_Panolis_flammea_NPV      AATAACGCCG GTTCCCAAAC TAATTGTTTC GGTGACAAAT CTAAAGAACG CGATGCCCGT CGTCGAATAC TCAGGTTTGG
K10_Sp_frugiperda_NPV       TACGACGCCG GTGCCCAAGT CTATAGTTTC GTTGACCAAT TTGAAAAACG CCATGCCTGT CGTCGCGTGC AATAATCTGG
K11Spodoptera_exempta_NPV   CATCACGCCG GTGCCGAAGT TGATTGTTTC GCTGACAAAC TTGAAAAATG CCATGCCCGT CGTGGCCTAT GATAATTTGC
K12_Malacosoma_neustria_NPV CACCGTGCCC GTGCCCAAAT TGATTGTGTC GTTGACAAAC TTGAAAAACG CAATGCCCAT ACTGGCCTAC GACGAAAGCG
K1lef8c                     TACCACGCCC GTACCCAAAT TGATTGTGTC CCTAACCAAT CTAAAGAATG CCATGCCCGT GTTTGAGTAC AAAGAGGAAA
K2lef8c                     TACCACGCCC GTACCCAAAT TGATTGTGTC TCTAACCAAT CTAAAGAATG CCATGCCCGT GTTTGAGTAC AAAAAAGAAA
K3lef8c                     CACCGTGCCC GTGCCCAAAT TGATTGTGTC GTTGACAAAC TTGAAAAACG CAATGCCCAT ACTGGCGTAC GATGAAAGCG
K8_sp_sunia_NPV             CACTATTCCC GTGCCGAAAC TCATCGTATC AGTGACGAAT CTGAAAAACG CGATGCCAGT GACCTGTTAC AACGAAGAAC
K9lef8c                     AATAACGCCG GTTCCCAAAC TAATCGTTTC GGTGACGAAC CTCAAAAACG CAATGCCCGT CGTCGAATAT TCAAGTTTGG
LdMNPV                      CACGCTGCCC GTGCCCAAGC TGATCGTGTC GCTGACCAAC TTGAAGAACG CGATGCCCGT GCAGAGCTAC ACGAAA----
OpMNPV                      CACCATTCCG GTGCCCAAAA TGATAGTGGC CCTGACAAAT CTGAAAAACG CGATGCCGGT GCTGCGGTAC AGCAGC----
PxGV                        CACCGTGCCG GTCGCCAAAC TGATCGTTAG TCTGACAAAT CTGAAAAACG GTATGGTGGT GCAGTGTCCG ACCAAAATAG
SeMNPV                      CACTACGCCG GTACCAAAGT CTATAGTTTC GCTGACGAAC CTGAAAAACG CCATGCCTGT CGTCACGTAC GACGGTTCGA
Sp_littoralis_NPV           GACGATACCC GTGCCCAAAC TGATCGTGTC CGTGACAAAT TTGAAAAACG CAATGCCCGT GATCGAGTAC GATAACGTCG
Splitura_NPVgenome          TACGATACCC GTGCCAAAAT TGATCGTGTC CGTGACAAAT TTGAAAAACG CAATGCCCGT AATCGAGTAC GACAACGTCG
XcGV                        GTCGGTTCCC GGAGCTAAAT TGATTGTGAG CCTGACAAAT CTTAAAAACG GTATGGTTGT CAATAGTCCG TTAAAATTAA
M2lef8c                     CACTACTCCT ATTGCAAAGT TAATTGTGAG TTTAACCAAT TTAAAAAACG GTATGTTGGT TCAAAAAAAC AAT-------
N1lef8c                     CACAATTCCG GTGCCCAAAA TGATAGGGGG CTTTNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
'P1lef8c=E1110'             CACACTCCGG GTGGCCAATT TAATGGTGGT CTTTTNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
CfMNPVlef8.txt              CACCATTCCG GTGCCCAAAA TGATAGTAGC CCTAACAAAC TTGAAAAACG CAATGCCGGT GTTCCGATAC AGCAGC----
Nele_NPV                    CAACATACCA CCGGCCAAAA TTATGGTATT TATGAGAAAT CTAAAAAACG CATCACCGGT ATTTATTCCA AGTAGCTGTG

10_Lasiocampa_quercus_NPV   ACGGG----- ---------- ------GTAT CGGGTTTACC AAATGGCTAC TCTGTGGCGG TAAACAAGTC C---ATTTTG
106_aporia_crataegi_NPV     ATTAT----- ---------- ------CTCG ACAACATTCC CGCTGGCTAT TCGGTCGGTG TGCACGAGTC C---ATATAT
112_Abraxas_grossulariata_NPV ---------- ---------- ------CTCG GCACGGTGCC GAGCGGGCTG TCGGTGGCGG TCGGCGACGG C---GTGCTA
126_Spilosoma_lutea_NPV     CTCCG----- ---------- ------TTTG ACCACCTGCC GCTGGGCACG AGCGTGAGTG TCGCACCGCA C---ATTATG
14_Anomis_sabulifera_NPV    TTACAAAAAG TAAATTA--- ------TTAT CAAATTTGCC GTACGGCAAT TCGGTGTTGG TTCGTCCCGA C---ATATGC
174_Erranis_defolaria_NPV   GCGACAAAGA CTTCTTC--- ------TTAG ACACATTGCC TCAGGGAAAT GCTGTAATAG TAGCGCCTCG A---ATCACA
175_Anomala_insulata_NPV    ---------- ---------- ------CGCG GCGCGCCGGC CGTCGGCTTG TCCGTGGCGG TCGGGGACGG C---GTGCTG
19_Junonia_coenia_GV        TCGAT----- ---------- ------CTCG ACGATCTACC GTCCGGCAAT AGCGCTTCCG TCTCCGACGA T---GTTTAC
193_Lymantria_obfuscata_NPV GCGCCGACG- ---------- ------CGGA CGCGCTGTCC TGTGGGCAAC TCGGTGGTGG TGTCGCCGCG C---GTGATG
195_Trabala_vishnou_NPV     ACGAACGCGT ACGA------ ------TTGG ACTGTTTGCC TCACGGCAAT TCGGTGGTGG TGGCGCCTCG C---ATTCAA
202_Lasiocampa_quercus_NPV  TCGACAACGA GCGCTTC--- ------CTAG ACACTCTGCC TTGCGGCTAT TCAGTCCTGG TCGATCCTGA G---CATAAG
205_Kotochalia_NPV          TT-------- ---------- ------ATCG ACGTACTTCC GTACGGCAAT TGTGTGGTGG TGACGCCGCG A---ATGAGG
228_Hyphantria_cunea_GV     AACTC----- ---------- ------TTAG ATTATTTACC TCTAGGCACT AGCGTGACCG TAAGTCCATC T---ATGATG
248_Euplasia_lucipara_GV    CGGAACAC-- ---------- ------GATC TGTTGTTACC CTTGGGTAAC AGTGTGTGTG TCGAAAAATC G---ATTTAC
254_Natada_nararia_GV       ---------- ---------- ------TTTA AAATTTTACC CATAGGTAAC AGTGTCACTG TAGACGAATC T---ATTTAT
26_Melanchra_persicae_GV    CGGAC----- ---------- ------TTTT CTTTTCTACC TATTGGTAAC AGTGTGGGTG TGAACGAGAG C---GTGTAT
274_Mamestra_brassicae_NPV  CGGACGCACA ACGCGCT--- ------TTGG ACAATTTGCC GTCCGGCAAT GCCGTGGTGG TGGCGCCCGA G---ATTATG
28_Samia_ricini_NPV         ACTGC----- ---------- ------GTAT CGGGTTTACC AAATGGCTAC TCTGTGGCGG TAAACAAGTC C---ATTTTG
291_O._brumata_NPV_dartmoor AA---CACCA CACGACG--- ------ATAG ACAACTTGCC TATTGGTAAT GTGATTGTGG TCGACCCCAA G---ATTCGA
30_Antherea_pernyi_NPV      ACTAC----- ---------- ------GTAT CGGGTTTACC AAATGGCTAC TCTGTGGCGG TAAACAAGTC C---ATTTTG
302lef8c                    ---------- ---------- ---------- --------CA GAACGGCAAA GGCGTGACCG GCTTGTGGAA A---GGATAT
32_Kotochalia_NPV           TT-------- ---------- ------ATCG ACGTACTTCC GTACGGCAAT TGTGTGGTGG TGACGCCGAA A---ATGAGG
342_Wiseana_GV              ---------- ---------- ------GTGG GAAATTTGCC TATGGGCAAT AGCGTTGTAG TAGACAAATC T---ATATAc
344_wiseana_cervinata_NPV   CGGAAAATCG CGCCGCTAGC ------CTCT TCGGGCTGCC CACCGGCAAC TTTGTCGTGG TCGACGACGC C---GCCTAC
353_Nymphalis_io_NPV        ATGATGGGGT TAATTTCGAA ACATTCCTCG ACACGTTACC TTTCGGCAAT TCTGTTGTCG TCACGCCCGC A---ATTAGA
362_Shinx_ligustri_NPV      ACTCC----- ---------- ------GTAT CGGGTTTACC AAATGGCTAC TCTGTGGCGG TAAACAAGTC C---ATTTTG
367_Vanessa_cardui_NPV      ATGATGGGGT TAATTTCGAA ACATTCCTCG ACACGTTACC TTTCGGCAAT TCTGTTGTCG TCACGCCCGC A---ATTAGA
37_cacoecia_murinana_GV     ---------- ---------- ------TATG AATTTTTGCC ATTGGGTAAC AGTGTGGTGG TGCATGATTC C---ATATAT
'380 Polygonia c-album NPV' ATGATGGGGT TAATTTCGAA ACATTCCTCG ACACGTTACC TTTCGGCAAT TCTGTTGTCG TCACGCCCGC A---ATTAGA
384_Pieris_brassicae_GV     ---------- ---------- ------TTTA AAATGCTTCC TCTTGGCAAT AGCGTTGTTG TTGACGATTC T---ATATAT
402_idea_virgularia_NPV     ---------- ---------- ------CTCG GGTTGGTGCC CGCCGGCCTA TCTGTAGCAG TTGGAGACAG C---GTTATC
436_Pachypasa_papyri_NPV    TTGAACAAAA TGATTTT--- ------TTGG ACACACTACC TGCAGGTTAT TCTCTTCTTG TTAACGCCGG C---ATTAAA
440lef8c                    AATGGACTGA AACGGGCAGT TCTTCATTAG AAGACTTGCC GGTCGGCAAT tCTGTCGTGG CGAGCCCCGG T---TCGGTG
```

460_Bombyx_mori_NPV GTTGC----- ---------- ------GTAT CGGGTTTACC AAACGGCTAC TCTGTGGCGG TGAACAAGTC C---ATTTTG
47_Actias_selene_NPV ---------- ---------- ------CGCG GCGCGCCGCC CGTAGGACTG TCCGTGGCGG TCGGCGACGG C---GTGCTG
505_buzura_edwardsiiNPV GCAACAACGA TATATTT--- ------TTGG ACAATTTACC GGCAGGCAAC AGCGTTGTCG TCGCTCCCGA G---ATTATG
54_Vanessa_atalanta_NPV ATGATGGCGC TAATTTTGAA ACATTCCTCG ACACGTTACC TTTCGGCAAT TCTGTTGTCG TCACGCCCGC A---ATTAGA
545_Darna_trima_GV ---------- ---------- ------AAGT TTACACCCGT AATTGGCAAC AGCGTTGTAG TGGACGAATC T---ATTTAT
595lef8c AGGTCTCTCA TGATGCGCAC ------GCTG ATGATGATGA TCATCATAAC CCTCAAATAT TCCATCATGA T---GATGTT
63_Aroa_discalis_NPV AACCGGCGTT T--------- ------GTGA ACGCGTTGCC GGCCGGCAAC TCGGTAATGG TGGCGCCGCG C---ATTCGT
64_Achaea_faber_NPV AAGACAACAA CGTGTTC--- ------TTGG ACAATCTGCC GATCGGCAAT TCCGTAGTGG TTTCGCCCGA C---ATTATG
66_sp_littoralis_GV TAAAT----- ---------- ------TTAG ACGAACTACC TGTGGGTAAT AGTGCTGCAG TAGATCCTTC T---ATAATG
67_Orgya_mixta_NPV AAGCGGCGTT T--------- ------GTGA ACGCGTTGCC GGCCGGCAAC TCGGTAATGG TGGCGCCGCG C---ATTCGT
692lef8c ATAACACTGA TGATGATGAT ------GATG ATGATGATGA TCATCATAAC CCTCAATTAT TCAATGATGA T---AATTTT
744lef8c ---------- ---------- ------AATA ACATTTTGCC ATTGGGTAGC AGTGTGGTGG TGGATGACTC A---ATTTAT
750lef8c NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN N---NNNNNN
766_Sp_litura_GV TAAAT----- ---------- ------TTAG ACGAACTACC TGTGGGTAAT AGTGTTGCAG TAGATCCTTC T---ATAATG
768_Euproctis_similis_NPV GACCT----- ---------- ------CTGT CCATGTTATC CGTTGGCAAT TCGGTAGGCG TGCACGATAG C---ATTTTG
78_Catopsila_pomona_NPV CCGTTGTC-- ---------- ------GCCG CCGCTCTGCC GAACGGATGT TCGGTGGCCG TGCACAAATC A---ATATTG
789_Heliconius_erato_NPV GC-------- ---------- ------GCGC TGCCGCCGCC CGTCGGGCTG TCCGTGGCGG TGAGCGACAC C---GTCATG
798_Diacrisia_obliqua_NPV --ACG----- ---------- ------TTAA AAGCGGTGCC CGCCGGCCTG TCCGTGGCGG TAAGCGACAA T---GTGTTC
A1_Mythmnia_unipunctata_NPV AAGACGAAAG CATCTTT--- ------CTCG ACAATCTGCC GATGGGCAAC GCAGTCGTCG TGTCGCCCGA G---ATTATG
a10_Euproctis_chrysorrohea_NPV GACCT----- ---------- ------CTGT CCATGTTATC CGTTGGCAAT TCGGTAGGCG TCCACGATAG C---ATTTTG
a2_Mythmnia_separata_NPV_Japan CGGACCGTGA AGTGTTT--- ------TTGG ACAATTTGCC GTCCGGCAAT GCGGTGGTGG TGGCGCCCGA G---ATTATG
a3_Mythmnia_separata_NPV_china GTAGCAAACG AACGTTC--- ------CTGA ACACGCTGCC GACGGGCAAC TCTGTGATCG TCGACGAACG G---ATCAGA
a4_Mythmnia_separata_NPV_India GTAGCAAACG AACGTTC--- ------CTGA ACACGCTGCC GACGGGCAAC TCTGTGATCG TCGACGAACG G---ATCAGA
a5_Cerapteryx_graminis_NPV AAAGCGCCGA AGCTTTC--- ------CTCG ACAATTTGCC CACCGGAAAC GCCGTGGTGG TGGCGCCTGA G---ATTATG
a6_Lacanobia_oleracea_GV NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN N---NNNNNN
a7_Aglais_urticae_NPV ACGACGGCGT TAATTTTGAA ACATTCCTCG ACACGTTACC TTTCGGCAAT TCTGTTGTCG TCACGCCCGC A---ATTAGA
a9_Orgya_antiqua_NPV AGAAAGCGTT T--------- ------GTGA ACGCGCTGCC GCAGGGCAAC TCGGTGGCGG TGGCGCCGCG C---ATTCGG
AcMNPV -----GAGGA AAACTGC--- ------GTAT CGGGTTTACC AAATGGCTAC TCTGTGGCGG TAAACAAGTC C---ATTTTG
B10_Anticarsia_gemnitalis GA-------- ---------- ------AATG CCGCGTTGCC CGTGGGCTTG TCAGTGGCGG TC-AGGACAC T---GTGTTA
b3_Plodia_interpunctella_GV ---------- ---------- ------TACA AAATTCTACC TATAGGAAAC AGCGTAATGG TGGGCGATAG T---GTTTAT
b4_Lymantria_monacha_NPV CGACGCAGCG ATCCGCCGTG TTT---TTGG AGGAGCTGCC CGCCGGCAAC TCGGTGGTGG TGGCGCCGCG C---ATAATG
b6_Orgya_leusostigma_NPV AATCGACGTT T--------- ------CTAA ACAAATTACC GCAAGGCAAT TCGGTGGTGG TGGCGCCTCG T---ATTAGA
BmNPV -----GAAAA TAGTTGC--- ------GTAT CGGGTTTACC AAACGGCTAC TCTGTGGCGG TGAACAAGTC C---ATTTTG
c2_Trichoplusia_ni_NPV ATTGC----- ---------- ------GTAT CGGGTTTACC AAATGGCTAC TCTGTGGCGG TAAACAAGTC C---ATTTTG
c3_Potato_tuber_GV_8 ---------- ---------- ------GGTA ATTTGCTACC GGTGGGAAAC AGTGTGCTCG TCGATGAATC A---ATCTAC
c4__Potato_tuber_GV_9 ---------- ---------- ------GGTA ATTTGCTACC GGTGGGAAAC AGTGTGCTCG TCGATGAATC A---ATCTAC
C5_Ephasia_cautella_NPV TCGACTACAA C--------- ------ATAA ACACGTTGTC TTACGGTTAC AGTGTAGCTG TCGCCGAATC C---ATCAAA
CpGV ---------- ---------- ------TTCA AGTCATTGCC GGTGGGCAAC AGTGTGGTGG TGGAGGACAG T---GTGTAT
242_pieris_napi_GV ---------- ---------- ------TTTA AAATGCTTCC TCTTGGCAAT AGCGTTGTTG TTGACGATTC T---ATATAT
95_pieris_rapae_GV ---------- ---------- ------TTTA AAATGCTTCC TCTTGGCAAT AGCGTTGTTG TTGACGATTC T---ATATAT
CuniNPVlef8 AGCCCAGTCC GCCC------ ------ACGT CAAGCATACC GGACCGGTAC GCGGTGGTGG TTTGGGGTGA T---AGTTAT
EppoNPV ---------- ---------- ------TTTA AAACGCCGCC AATCGGCCTG TCCGTGGTAG TTAGCGACAA C---GTGCTC
HaSNPV GC---GATCA CAACGAA--- ------TTGA ATATGTTGCC CAATGGCAAT TCGGTTGTGG TGGCACCTCA C---ATTTTG
HzSNPV GC---GATCA CAACGAA--- ------TTGA ATATGTTGCC CAATGGCAAT TCGGTTGTGG TGGCACCTCA C---ATTTTG
i3_Panolis_flammea_NPV CGAACAAACA ACGTCTT--- ------TTGG ACAATCTACC GTCCGGCAAC GCGGTGGTAG TGGCGCCAGA T---ATTATG
K10_Sp_frugiperda_NPV GAACAAACA TATATTC--- ------TTGG ACAATTTACC GGCAGGCAAC AGCGTTGCCG TCGCTCCCGA G---ATTATG
K11Spodoptera_exempta_NPV AACACAACAA CGTGTTC--- ------TTGG ACAATCTGCC GATCGGCAAT TCCGTAGTGG TTTCGCCCGA C---ATTATG
K12_Malacosoma_neustria_NPV CCAATCGACA ACTCTTC--- ------TTGG ACACTTTGCC GACCGGCTAC TCGATGATCG TCAGTCCCGA T---ATCAAA
K1lef8c ATTGC----- ---------- ------GTAT CGGGTTTACG AAATGGCTAC TCTGTGGCGG TAAACAAGTC C---ATTTNN
K2lef8c ATTGC----- ---------- ------GTAT CGGGTTTACC AAATGGTTAC TCTGTGGCGG TAAACAAGTC C---GTTTTG
K3lef8c CCGATCGACA ACTCTTC--- ------TTGG ACACTTTGCC GACCGGCTAC TTGATGATCG TCAGTCCCGA T---ATCAAA
K8_sp_sunia_NPV TACAACTTAA TGTTTTT--- ------TTAA AGACTTTGCC GGTGGGTAAT TCGGTGGTGG TGAGTCCTCG A---ATACAA
K9lef8c CCGACCGACA AGTGTCT--- ------TTGG AAAATTTGCC GTCCGGCAAT GCGGTGGTGG TGGCGCCCGA G---ATTATG
LdMNPV -----GAGGG GGTCGAC--- ------GCGG ACGCGCTGCC CGTGGGCAAC TCGGTGGTGG TGGCGCCGCG C---GTGATG
OpMNPV ---------- ---------- ------CTGG GCACGGCGCC CGTCGGGCTG TCCGTGGCGG TCGGCGACCA T---GTGTTG
PxGV ACAAT----- ---------- ------TACG ATTATTTACC GCTGGGCAAC AGCGTCAGCG TCGATCCGAA A---TTAATG
SeMNPV GC---GAGGC GAAGCGG--- ---TTTCTCG ACAATCTGCC CGTGGGCAAT AGCGTCGTCG TCGCTCCCGA G---ATTATG
Sp_littoralis_NPV AATCGAACGA AAACCTCTGAC ATCGTCCATTA GAAACTTGCC GGTCGGCAAT TCTGTCGTGG CGAGCCCCGG T---TCGGTG
Splitura_NPVgenome AATCGAACGA GAACCCCGAC ATCGTCCATTA GAAACCTGCC TGTCGGCAAT TCCGTGGTGG CGAGCCGTAG C---TCGGTT
XcGV AAAAT----- ---------- ------TTAG ACTGGTTACC TTTAGGTCAC AGCGTTGCCG TGGACGATTC G---ATTCGT
M2lef8c ---------- ---------- ------TTTT CATTTTTACC CCTTGGAAAT AGTGTAATAA TCAACGACGA CAACATTTAT
N1lef8c NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
'P1lef8c=E1110' NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
CfMNPVlef8.txt ---------- ---------- ------CTTG GAACGGTGCC CGCCGGACTG TCCGTGGCGG TCAGCGACGG C---GTGATG
Nele_NPV TAGTTGAAAA ATATATGAAA ------TTTA TGAAAAATGA AAATAAAAAC TGCATTGTTG TGGATTCGAA A---CATTAT

10_Lasiocampa_quercus_NPV CTCAACAATA AAATGNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNN [684]
106_aporia_crataegi_NPV ACTAACAACA AAATGTTCAA ATTATGGACG CTGGTTCGCG ACAATAGACT GATGACA [687]
112_Abraxas_grossulariata_NPV ATGAACAACA AGATGGTCAA GCTGTGGACG CTGGTGCGAG ATTCCAAGCT CATGACG [684]
126_Spilosoma_lutea_NPV TGCAACAATC GCATGTTCAA TTTGTGGACT TTGGTGCGTG ATAACAAATT AAAAACC [792]
14_Anomis_sabulifera_NPV ATAAACAACA AAATGTTTAA ACTGTGGACA CTGTTGCGAG ATAACAAATT GAAAACG [687]
174_Erranis_defolaria_NPV TTGAACAACA AAATGTTCAA ACTGTGGACA ATAGTTCGAG ATCACCGATT GATGACC [727]
175_Anomala_insulata_NPV CTAAACAACA AGATGGTGCG GCTGTGGACG TTGGTGCGCG ACTCGCGGCT CATGACC [684]
19_Junonia_coenia_GV TTCAACGATC GCATGTTTTG TCTGTGGACG ATCGTGCGAG ACAATCGCcT GAAACCG [663]
193_Lymantria_obfuscata_NPV CTCACAATAA AATGTTTNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNN [946]
195_Trabala_vishnou_NPV CTCAACAACA AAATGTTCAA GCTGTGGACG TTGGTGCGCG ATCACAAACT CAGACCG [795]
202_Lasiocampa_quercus_NPV CTAAACAACA AAATGTTTAA ATTGTGGACA CTGGTGAGGG ACTACATACT CATGACG [698]
205_Kotochalia_NPV ATTAACAATA AAATGTTTAA ACNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNN [669]
228_Hyphantria_cunea_GV TGCAACGACC GCATGTTTTA TCTGTGGACT CTAGTGCGCG ATAATAAACT GAAAACG [789]
248_Euplasia_lucipara_GV TACAACGATC GTATGTTTAA TCTGTGGACC CTGGGCGCGA TAACAGACTC AGACCGC [943]
254_Natada_nararia_GV TGTAACGATC GGATGTTTTA TTTGTGGACA TTAGTTAGGG ATAATAAATT AAAAACG [678]
26_Melanchra_persicae_GV TACAACGATC GCATGTTTAA TTTGTGGACT ATAATTCGAG CCAACAAATT AAAGCCG [822]
274_Mamestra_brassicae_NPV CGCAACGACA AAATGTTCTA TTTGTGGACG ATTGTGCGCG ACNNNNNNNN NNNNNNN [702]
28_Samia_ricini_NPV CTCACACAAa AGAATGTTaA AATGTGGACN NNNNNNNNNN NNNNNNNNNN NNNNNNN [685]
291_O._brumata_NPV_dartmoor GTCAACAACA AAATGTTTAA ACTGTGGACG TTGATACGTG ACTTTAGACT CAAGACG [684]
30_Antherea_pernyi_NPV CTCAACAATA AAATGTTTAA ATTGTGGACG TTGGTGCGAG ACAATAAACT CATGACC [686]
302lef8c -TAGGTNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNN [920]
32_Kotochalia_NPV ATTCACAATA AAATGTTTAA ACTTTNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNN [669]
342_Wiseana_GV TGCAACGTCC GTATGTTTTA CTTGTGGGCT CGGgTGAGGA ANNNNNNNNN NNNNNNN [671]

344_wiseana_cervinata_NPV	CTCACCAACA AAACCTTCAA AATGTGGACC GTCATCGCGG ACCGCCGCCT TGGCACC	[737]
353_Nymphalis_io_NPV	CAAAATAATA AAATGTGCCA ATTGTGGACG TTGGTGCGCG ACAACAAACT CATGACG	[732]
362_Shinx_ligustri_NPV	CTCAACAATA AAATGTTTAA ATTGTGGACG TTGGTGCGAG ACAATAACTC ATGACCG	[684]
367_Vanessa_cardui_NPV	CAAAATACTN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNN	[733]
37_cacoecia_murinana_GV	TACAACCATC GCGCTTTTCG TATTATGGGC CACCNNNNNN NNNNNNNNNN NNNNNNN	[729]
'380 Polygonia c-album NPV'	CAAAATAATA AAATGTGCAA ATTGTGGACG TTGGTGCGCG ACAACAAACT CATGACG	[732]
384_Pieris_brassicae_GV	TACAACGATC GCATGTTTTG TTTGTGGACA ATAGTTATGG ATAAAAAACT AAAAACG	[771]
402_idea_virgularia_NPV	ATGAACAACA AAATGCTCAA GTTGTGGACG CTAGNNNNNN NNNNNNNNNN NNNNNNN	[684]
436_Pachypasa_papyri_NPV	CTCAATAACA AAATGTTTAA ATTGTGGACA CTTGTGAGAG ATTTTAAATT AATGACG	[720]
440lef8c	CGCAACAGCA AGATGNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNN	[960]
460_Bombyx_mori_NPV	CTCAACAATA AAATGTTTAA ATTGTGGACG TTGGTGCGAG ACAATAAACT CATGACC	[684]
47_Actias_selene_NPV	CTAAACAACA AAATGGTGCG GCTGTGGACG CTGGTGCGCG ACTCAAAAAC TCATGAC	[684]
505_buzura_edwardsiiNPV	CGCAACGACA AAATGTTTCG TTTGTGGACA CTGGTGCGCG ATCATAAACT GATGACC	[693]
54_Vanessa_atalanta_NPV	CAAAATAATA AAATGTGCAA ATTGTGGACG TTGGTGCGCG ACAACAAACT CATGACG	[732]
545_Darna_trima_GV	TGTAACGATC GCATGTTCTA TTTGTGGACG GTCGTAATGG ATAGCGCTCT TAAGACG	[675]
595lef8c	TTAAATAATT tACCACAGGG TAATTCGGTG GTCGTGGCGC CCCCAATCAT CTTAATG	[972]
63_Aroa_discalis_NPV	CTCAACAACA AAATGTTCAA CTTGTGGACG CTGGTGCGCG ACCACCGACT CATGACC	[711]
64_Achaea_faber_NPV	CGTAACGACA AAATGTTTCG TCTGTGGACG TTGGTGCGCG ACCACAAACT CATGACC	[702]
66_sp_littoralis_GV	TGCCACGATC GTATGTTTCG ACTGTGGTCC ATAATTCGAG ATTNNNNNNN NNNNNNN	[660]
67_Orgya_mixta_NPV	CTCAACAACA AAATGTTCAA CTTGTGGACG CTGGTGCGCG ACCACCGACT CATGACC	[711]
692lef8c	TTAAATAATT TACCACAGGG TAATTCGGTG GTCGTGGCAC CTGAAATCAT TCTTAAT	[711]
744lef8c	TGCAACGATC GGCACTTTTG NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNN	[742]
750lef8c	NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNN	[697]
766_Sp_litura_GV	AGCAACGATC GTATGTTTCG ACTGTGGNNN NNNNNNNNNN NNNNNNNNNN NNNNNNN	[660]
768_Euproctis_similis_NPV	ATGAACAACA AGATGTTTAA ACTGTGGACT TTGGTGCGCG ACAACAGGCT GATGACG	[687]
78_Catopsila_pomona_NPV	ATGAACAACA AAATGTTCAA ACTGTGGACG TTAGTGAGAG ACAACCGTTT GAGACTG	[696]
789_Heliconius_erato_NPV	ATGAACAACA AAATGGTCAA ATTGTGGACG CTGGTGCGCG ACTCGAAGCT CATGACC	[702]
798_Diacrisia_obliqua_NPV	ATGAACAACA AAATGGTGAA GCTGTGGACG CTCGTGCGCG ATTCAAAACT CAGACCG	[684]
A1_Mythmnia_unipunctata_NPV	CGAAACGACA AAATGTTCAG ACTGTGGACG CTGGTGCGAG ACAATAGACT GCAAACC	[717]
a10_Euproctis_chrysorrohea_NPV	ATGAACAACA AAATGTTTAA ACTGTGGACT TTAGTGCGCG ACAACAAGCT GATGACG	[687]
a2_Mythmnia_separata_NPV_Japan	CGCAACGACA AAATGTTCTA TTTGTGGACG ATTGTGCGCG ACAACAAGCT ACAAACC	[702]
a3_Mythmnia_separata_NPV_china	CAGAACAACA AAATGTTCAA ACTGTGGACG CTGGTGCGCG ACAACAGCCT CTACACC	[721]
a4_Mythmnia_separata_NPV_India	CAGAACAACA AAATGTTCAA ACTGTGGACG CTGGTGCGCG ACAACAGCCT CTACACC	[720]
a5_Cerapteryx_graminis_NPV	CGAAACGACA AAATGTTTAG GTTATGGACG CTGGTGCGCG ACAACAAACT GATGACC	[696]
a6_Lacanobia_oleracea_GV	NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNN	[845]
a7_Aglais_urticae_NPV	CAAAATAATA AAATGTGCAA ATTGTGGACG TTGGTGCGCG ACAACAAACT CATGACG	[732]
a9_Orgya_antiqua_NPV	TTCAATGATA AAATGTTTAA ACTGTGGACG CTGGTGCGCG ACCACCGCCT CATGACG	[714]
AcMNPV	CTCAACAATA AAATGTTTAA ATTGTGGACG TTGGTGCGAG ACAATAAACT CATGACC	[684]
B10_Anticarsia_gemnitalis	ATGAACAACA AAAGGGGTAA ACTGTGGACT TTGGAGCGCG ATTTTAAACC NNNNNNN	[690]
b3_Plodia_interpunctella_GV	TACAACGATA GAATGTTTTC GCTGTGGACG TTGGTGAGGG ATAACAAATT GAAAACC	[723]
b4_Lymantria_monacha_NPV	GTCAACAACA AGATGTTCAA GCTGTGGACG CTGGTGCGCG ACCATCGATT GATGACC	[732]
b6_Orgya_leusostigma_NPV	CTCAACAACA AAATGTTCAA TCTTTGGACG TTGGTGCGAG ACCACCGGCT CATGACG	[706]
BmNPV	CTCAATAATA AAATGTTTAA ATTGTGGACG TTGGTGCGAG ACAATAAACT CATGACC	[684]
c2_Trichoplusia_ni_NPV	CTCAACAATA AAATGTTTAA ATTGTGGACG TTGGTGCGAG ACAATAAACT CATGACC	[684]
c3_Potato_tuber_GV_8	TGTAACGACC GCATGTTTTA TTTGTGGACT CTGGTCATGG ATATCACCTT GAAAACG	[654]
c4__Potato_tuber_GV_9	TGTAACGACC GCATGTTTTA TTTGTGGACT CTGGTCATGG ATATCAAATT GAAAACG	[654]
C5_Ephasia_cautella_NPV	ACCAATAACA AAATGTTCAA ATTGTGGACT TTGGTGCGAG ACAACAAACT TATGACC	[690]
CpGV	TGTAACGATC GCATGTTTTG TTTGTGGACG GTGGTGATGG ACAGTAGGTT GAAGACG	[756]
242_pieris_napi_GV	TACAACGATC GCATGTTTTG TTTGTGGACA ATAGTTATGG ATAAAAAACT AAAAACG	[771]
95_pieris_rapae_GV	TACAACGATC GCATGTTTTG TTTGTGGGCA ATAGTTATGG ATCTGGNNNN NNNNNNN	[771]
CuniNPVlef8	CTCAACGAGG AATCGTTTGC GTTGTGGACA ATTGTGGGAG ATTTTCACAA CGGAACC	[693]
EppoNPV	ATGAACAATA AAATGGTAAA ATTGTGGACA CTGGTGCGCG ATTCTAAACT GATGACG	[684]
HaSNPV	CACAACAACA AAATGTTCAA ACTTTGGACG TTGGTACGTG ACGATCGATT ACAGACC	[759]
HzSNPV	CACAACAACA AAATGTTCAA ACTTTGGACG TTGGTACGTG ACGATCGATT ACAGACC	[759]
i3_Panolis_flammea_NPV	CGCAACGACA AAATGTTCTA TCTGTGGACC ATTGTGCGCG ACAACAAACT GCAAACG	[702]
K10_Sp_frugiperda_NPV	CGCAACGACG AATGTTTCGT CTTGTGGACA CTGGTGCGCG ATCATAGACT GATGACC	[693]
K11Spodoptera_exempta_NPV	CGTAACGACA AAATGTTTCG TCTGTGGACG TTGGTGCGCG ACCACAAACT CATGACC	[702]
K12_Malacosoma_neustria_NPV	TTGAACAACA AAATGTTCAA ACTGTGGACG CTGGTGAGGG ACTTTAAGTT GATGACC	[699]
K1lef8c	NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNN	[685]
K2lef8c	CTCCACAATA AAATGTTTAA ATTGTGNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNN	[684]
K3lef8c	TTGAACAACA AAATGTTCAA ACTGTGGACG CTGGTGAGGN NNNNNNNNNN NNNNNNN	[700]
K8_sp_sunia_NPV	GTGAATAATA AAATGTGCAA ACTCTGGAcA ATTGTACGAG -CCTAATAAN NNNNNNN	[689]
K9lef8c	CGCAACGACA AAATGTTCTA TCCTGTGNNN NNNNNNNNNN NNNNNNNNNN NNNNNNN	[701]
LdMNPV	CTCAACAATA AAATGTTCAC CCTGTGGACG CTGGTGCGCG ACCACCGGCT CATGACC	[687]
OpMNPV	ATGAACAACA AGATGGTCAA GTTGTGGACG CTGGTGCGCG ACTCCAAGCT CATGACC	[684]
PxGV	TTTAATAATC GCATGTTTTA TCTTTGGACT TTGGTTAGAG ATAGCAATTT AAAAACG	[672]
SeMNPV	CGCAACGACA AAATGTTTCG GCTGTGGACG CTGGTGCGCG ACCACAAGCT CATGACC	[744]
Sp_littoralis_NPV	CACAACAACA AGATGATCAA TCTGTGGACT CTGGTGCGAG ACAGCCGGCT CATGACC	[717]
Splitura_NPVgenome	CACAACAATA AAATGATAAA TTTGTGGACT CTGGTGCGAG ACAGCCGACT CATGACC	[744]
XcGV	TGTAACGATA GAATGTTTTG TCTGTGGACC ATTGTGCGCG ATAGCAAATT AAAAACA	[681]
M2lef8c	TGTAATGACC GAATGTTTTA TTTATGGACA ATAGTGTTGG ACAAAAATAA AANNNNN	[690]
N1lef8c	NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNN	[693]
'P1lef8c=E1110'	NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNN	[682]
CfMNPVlef8.txt	ATGAACAACA AAATGGTTAG GCTGTGGACG TTGGTGCGCG ATTCTAAACT CATGACC	[684]
Nele_NPV	TGTAACGACA AAATGATATA CTTGTGGACA ATGATCGGAG ATGATAGATT AAAAACT	[651]