

Université du Québec
Institut National de la Recherche Scientifique
Institut Armand-Frappier

**New densoviruses, comparison of transcription strategies,
recombination and contribution to 3D-structures**

Par
Qian Yu

Thèse présentée pour l'obtention du grade de
Philosophiae doctor (Ph.D.) en Virologie et Immunologie

Jury d'évaluation

Président du jury et
Examinateur interne

Jean-François Laliberté
INRS-Institut Armand-Frappier

Examinateur interne

Pierre Talbot
INRS-Institut Armand-Frappier

Examinateuse externe

Selena M. Sagan
Department of Microbiology and
Immunology
Mcgill University

Examinateur externe

Just M. Vlak
Laboratory of Virology,
Wageningen University

Directeur de recherche

Peter Tijssen
INRS-Institut Armand-Frappier

ACKNOWLEDGEMENTS

I would like to thank Prof. Peter Tijssen for giving me the opportunity to study at INRS-Institut Armand Frappier, and providing me precious advice and direction throughout my PhD study. He has created a warm and stimulating research environment allowing me explore a complex and fascinating scientific field. Thanks to him for his kindness, patience and support during my PhD path. I learned lots from him. Being able to study under his supervision is one of the most important achievements in my life.

I want to acknowledge Dr. Kaiyu Liu. He gave me invaluable help at the beginning of my PhD. I am grateful to Prof. Max Bergoin from the “Laboratoire de Pathologie Comparée, Université Montpellier II, France” and now as an invited professor in INRS- Institut Armand-Frappier for his important comments during my project research. Thanks to all the group members for their help throughout the work. I am grateful to have been part of such supportive, hardworking, and inspiring groups. Over the course of my research I have been fortunate enough to work with the truly outstanding groups. These people include: Dr. Maude Boisvert, Dr. Hanh Thi Pham, Dr. Sandra Fernandes, and Oanh T. H. Huynh.

I thank to my PhD committee (comité d'encadrement): Dr. Jean-François Laliberté and Dr. Jonathan Perreault, both professors at INRS-Institut Armand Frappier, for their kindly discussions and advice.

I thank the departmental and technical staff at INRS- Institut Armand-Frappier. They are very helpful, particularly Micheline Letarte for her EM technical support and Jessy Tremblay for his microscopy technical support.

Sepcial thanks to Madame Anne Philippon from the administration unit of the INRS-Institut Armand-Frappier, who always help me with many patience for the applications and document works.

I am eternally grateful to my dearest parents, who have been a constant source of love, encouragement for me. I want to thank to my dear husband, Jianming Zhang, who loves me deeply and supports me selflessly. I want to say them, I love you. I thank my family and my friends for their continuing love and heartily support.

Finally, I wish to acknowledge the following organizations for their financial support: the Chinese Scholarship Council and Natural Sciences and Engineering Research Council of Canada.

ABSTRACT

Parvoviruses are small, isometric viruses, with a diameter of 20–25 nm, containing a linear, single-stranded DNA genome of 4–6 kb, and have been classified in the family *Parvoviridae*. Densovirus (DNVs) are parvoviruses of invertebrates, infecting at least insects, shrimp, and urchins, and form a separate subfamily (*Densovirinae*) within the *Parvoviridae* family. The nonenveloped icosahedral capsids consist of 60 structural proteins some of which have N-terminal extensions. Moreover, capsids of many vertebrate parvoviruses and DNVs have a phospholipase A2 (PLA2) activity that allows them to breach the endosomal membrane barrier during cell entry. During the past few decades, many new vertebrate parvoviruses and invertebrate densovirus have been discovered. However, research on their molecular biology characteristics is still limited.

This thesis contains three different projects. The first objective, which is also the major project of this thesis, started with the discovery of three new iteradenoviruses. Since 2010, we got some infected *Papilio polyxenes* (black swallowtail butterfly) larvae and some infected monarch butterfly pupae from a lab of Cornell University and a farm in Granby (Quebec), respectively. Together with an older, stored *Sibine fusca* larvae sample, we identified three new viruses. The method we used for identification was named DNase&RNase-SISPA. Cloning and sequencing of the genomes of those three viruses revealed high identities with members of *Iteradenovirus* genus of *Densovirinae* subfamily. These new iteradenoviruses have later been named as *Papilio polyxenes* iteradenovirus (PpDV), *Sibine fusca* iteradenovirus (SfDV) and *Danaus plexippus plexippus* iteradenovirus (DppIDV).

Together with the other two iteradenovirus *Casphalia extranea* iteradenovirus (CeDV) and *Bombyx mori* iteradenovirus (BmDV) which have been cloned and reported from our lab, in the second project we determined the transcription strategies of these five iteradenoviruses. We showed that all these monosense

iteradenoviruses genomes contain two large reading frames (ORFs) coding for two non-structural (NS) proteins. NS1 and NS2 that are generated from individual transcripts, which initiated from either side of NS1 ATG with only 10-15 nt distance between each other. The expression of the two NS proteins is controlled by two overlapping promoters. The two NS transcripts shared the same poly(A) motif downstream of the TATA-box of VP. All the structural proteins were generated from a single VP transcript by leaky scanning.

The third project contained different studies on cricket densoviruses, which included a contribution work on the transcription strategy of AdDNV and the discovery of two novel cricket densoviruses. AdDNV belongs to the *Ambidensovirus* genus and its expression strategy of NS proteins is identical with the classical ambidensoviruses. However, the structural proteins (VPs) were generated from two split ORFs and employed a unique alternative splicing mechanism for expression. So far, this expression strategy is not only unique among the other densoviruses, but also among the vertebrate parvoviruses. The first novel cricket densovirus reported in this project was *Acheta domesticus* Mini Ambidensovirus AdMADV. The novel AdMADV has a small ambisense genome of 4945 nts and the related 199 nt inverted terminal repeats (ITRs) fold into Y-shaped hairpin structure at both telomeres. However, its genome organization is identical to classical ambidensoviruses. Another novel cricket densovirus consisted of a special segmented genome, the segments of which coded separately for the NS and VP proteins.. This new *Acheta domesticus* segmented densovirus, AdSDV, as the first bipartite parvovirus has been reported, arose from mosquitoes and through recombination with AdDNV to obtained phospholipase A2 activity that enabled it to infect crickets. Here, we also provided a hypothesis on the origin of this new virus.

Besides the main chapters for the three projects, the thesis also includes some published side projects in the appendix. The first project included in the appendix I aimed to determine the capsid structure for AdDNV and an adeno-associated virus

isolated from snakes (serpine adeno-associated virus, sAAV). For AdDNV, the 3.5-Å resolution X-ray crystal structure of the mature virus particle has been determined. It was shown that raising the temperature of the AdDNV particles caused a loss of the ssDNA genome. A 5.5-Å resolution structure of the empty virion was obtained by cryoelectron microscopy and was found to be the same as that for the full, mature virus except for the absence of the three ordered nucleotides. So far, serpine AAV (sAAV) is the only known parvovirus to infect reptiles (although recently, we got the complete sequence of a second reptile AAV from bearded dragon lizards (*Pogona*), unpublished). In this study, we determined the capsid structure at 6.7-Å resolution. Compared to reported structures of different serotypes of AAVs, significant differences were caused by the conformational changes within the variable regions VR-IV, VR-VIII, and VR-IX. Structural comparisons show that vertebrate and invertebrate parvoviruses have evolved independently, although there are common structural features among all parvovirus capsid proteins. Appendix II contains publications concerning the genome sequences of three novel single-stranded DNA (ssDNA) viruses. These viruses include the *Helicoverpa armigera* Densovirus HaDNV, *Pseudoplusia includens* Densovirus PiDNV, and a novel Circo-Like virus isolated from *Penaeus monodon* shrimp (PmCV-1). PmCV-1 was isolated from PstDNV-infected shrimps, *Penaeus monodon*, and was the first circular ssDNA virus that has been identified in shrimp.

RÉSUMÉ

Les Parvovirus sont de petit virus icosaédriques d'un diamètre d'environ 20-25nm, contenant un génome d'ADN simple brin de 4-6kb et sont classifiés dans la famille *Parvoviridae*. Les Densovirus (DNVs) sont des parvovirus infectant les invertébrés, incluant des insectes, les crevettes et les oursins et forme la sous-famille *Densovirinae* dans la famille des *Parvoviridae*. La capsid des parvovirus contient 60 protéines structurales dont certaines contiennent une extension N-terminale par rapport aux autres. De plus, la capsid de plusieurs parvovirus des vertébrés et DNVs possède une activité enzymatique de phospholipase A2 (PLA2) qui permet l'évasion de la voie endosomale lors de l'entrée du virus dans la cellule, via des bris de la membrane endosomale. Lors des dernières décennies, plusieurs nouveaux parvovirus de vertébrés et invertébrés ont été découverts. Cependant les connaissances concernant leur biologie moléculaire demeurent limitées.

Cette thèse contient trois projets différents. Le premier objectif, qui était le projet majeur de cette thèse a débuté par la découverte de trois nouveaux Iteradensovirus. Depuis 2010, nous avons reçu des échantillons de larves de papillons du céleri (black swallowtail butterfly, *Papilio polyxenes*) d'une ferme de Granby ainsi que des pupes de papillons monarques provenant de l'Université Cornell et Granby (Québec). Nous avions aussi en notre possession des échantillons de larves de *Sibine fusca*. À partir de ces échantillons, nous avons isolé et identifié trois nouveaux virus. La méthode utilisée pour leur identification est nommée « DNase&RNase-SISPA ». Le clonage et le séquençage des génomes de ces trois virus ont révélé une forte identité avec les membres du genre *Iteradensovirus* de la sous famille *Densovirinae*. Ces nouveaux *Iteradensovirus* ont été nommés *Papilio polyxenes* iteradensovirus (PpDV), *Sibine fusca* iteradensovirus (SfDV) et *Danaus plexippus* plexippus iteradensovirus (DppIDV).

Avec deux autres Iteradensovirus qui ont été clonés précédemment dans notre laboratoire : *Casphalia extranea* iteradensovirus (CeDV) and *Bombyx mori*

iteradensovirus (BmDV), la stratégie de transcription de ces 5 iteradensovirus est déterminée dans le second projet. Nous avons montré que ces virus sont tous monosens et contiennent deux grands cadres de lectures (ORF). Le premier permet l'expression de deux protéines non structurales (NS1 et NS2) qui sont générées par des transcrits indépendants, initiés à partir de séquences ATG à seulement 10-15 nucléotides de distance. L'expression des protéines NS est contrôlée par des promoteurs chevauchants. Les deux transcrits NS utilisent le même signal de polyadénylation (poly(A)), situé en aval du motif TATA-box des gènes des protéines structurales (VPs). Les protéines VPs sont toutes générées à partir d'un seul transcrit, par un mécanisme de « leaky scanning » (traduction par balayage fuite des codons d'initiation).

Le troisième projet contient différents études sur les densovirus de criquets, incluant un travail de contribution sur la stratégie de transcription d'AdDNV et la découverte de deux nouveaux densovirus de criquets. AdDNV appartient au genre *Amibendenvirus* et sa stratégie de transcription des protéines NS est identique aux autres membres de ce genre. Cependant, la stratégie de transcription des protéines VPs est unique. Les protéines structurales sont générées à partir de deux cadres de lecture fendues et emploie une stratégie d'épissage alternatif unique. Cette stratégie est non seulement unique au sein des densovirus mais des parvoviruses de vertébrés aussi. Le premier nouvel densovirus de criquet rapporté de ce projet était *Acheta domesticus* Mini Ambidensovirus AdMADV. Le nouvel AdMADV possède un petit génome de 4945 nucléotides ainsi que des séquences terminales inversées (ITR) de 199 nucléotides, repliées en forme de Y aux deux extrémités. L'organisation génomique est similaire par rapport aux autres ambidensovirus classiques. Un autre nouvel densovirus de criquet contient une séquence nucléotidique spéciale, où les deux séquences ont codé les protéines de NS et VP séparément. La nouvelle séquence de densovirus AdSDV de *Acheta domesticus* qui est le premier parvovirus en deux parties, a été rapportée d'émerger de moustiques et par la recombinaison avec AdDNV obtenu d'une activité de la phospholipase A2 qui était capable de infecter

les criquets. En plus, une hypothèse de l'origine de ce nouvel virus est aussi présentée ici.

Sauf les chapitres principaux de ces trois projets, on a aussi conclu des travaux de collaboration comme les projets secondaires dans l'annexe. Le premier projet inclus dans l'annexe I consistait à la détermination de la structure du densovirus AdDNV et d'un parvovirus isolés à partir de serpents (Serpine adeno-associated virus (sAAV)). Pour AdDNV, la structure de la particule mature a été déterminée par cristallographie à rayons X avec une résolution de 3.5-Å. Nous avons aussi montré que l'augmentation de la température des particules menait à l'éjection des génomes. Une structure du virion vide de résolution de 5.5 Å a été obtenue par microscopie cryo-électronique et est identique à la structure de la capsid pleine, excepté pour les trois nucléotides ordonnés présent dans la structure de la capsid pleine. sAAV est le seul parvovirus de reptile connu (bien que ces dernières années nous avons obtenu la séquence complète d'une seconde reptile AAV à partir de lézards dragons (*Pogona*), non publié). Nous avons déterminé sa structure à une résolution de 6.7-Å. En comparant cette structure avec les autres AAVs connus, certaines différences significatives ont été observées. Elles sont dues à un changement de conformation dans les régions variables VR-IV, VR-VIII et VR-IX. La comparaison des structures a montré que l'évolution des parvovirus des vertébrés et invertébrés a été effectuée indépendamment, malgré le fait que certaines caractéristiques structurales soient partagées chez tous les parvovirus. L'annexe II a contenu la séquence nucléotidique de trois nouveaux virus à ADN simple brin. Ces virus incluent *Helicoverpa armigera* Densovirus (HaDNV); *Pseudoplusia includens* Densovirus (PiDNV); et un nouveau virus à génome circulaire (Circo-Like) isolée à partir de crevettes *Penaeus monodon* (PmCV-1). PmCV-1, isolées à partir de crevettes infectées par le virus PstDNV est le premier virus avec un génome d'ADN simple brin circulaire identifié chez les crevettes.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS.....	ii
ABSTRACT.....	iv
RÉSUMÉ.....	viii
TABLE OF CONTENTS.....	xii
LIST OF TABLES.....	xviii
LIST OF FIGURES.....	xx
LIST OF ABBREVIATIONS.....	xxii
Chapter I: Introduction	2
1 Viruses and biological control	4
2 Insect parvoviruses (Densoviruses)	6
2.1 Genotype and Taxonomic structure history of Densovirus	6
2.2 Pathology caused by densoviruses	10
2.3 Genome organization of densoviruses	10
2.3.1 Genome organization of the <i>Ambidensovirus</i> genus.....	12
2.3.2 Genome organization of the <i>Iteradensovirus</i> genus.....	15
2.3.3 Genome organization of the <i>Brevidensovirus</i> genus.....	16
2.3.4 Genome organization of the <i>Hepandensovirus</i> and <i>Penstyldensovirus</i> genus	16
2.4 Transcription strategy of densoviruses	17
2.5 Terminal structures of densovirus genome	24
2.6 Viral proteins	27

2.6.1	Nonstructural protein and rolling cycle replication	27
2.6.1.1	Rolling hairpin replication	28
2.6.2	Structure protein	32
3	Atomic structures of viral particles	35
Chapter II: Materials and Methods		42
Chapter III: New iteradensoviruses.....		52
Preamble.....		54
Article 1: Iteravirus-Like Genome Organization of a Densovirus from <i>Sibine fusca</i>		
Stoll		56
Contribution of authors.....		56
Abstract.....		56
Résumé.....		56
Results.....		57
Achnowledgments.....		59
Article 2: <i>Papilio polyxenes</i> Densovirus Has an Iteravirus-Like Genome Organization..		60
Contribution of authors.....		60
Abstract.....		60
Résumé.....		60
Results.....		61
Achnowledgments.....		62
Article 3: Iteradensovirus from the Monarch Butterfly, <i>Danaus plexippus plexippus</i> ...		64
Contribution of authors.....		64
Abstract.....		64
Résumé.....		64
Results.....		65
Achnowledgments.....		67
Chapter IV: Iteradensovirus transcription strategy.....		68
Preamble.....		70

Article 4: Gene expression of five different iteradensoviruses: BmDV, CeDV, PpDV, SfDV and DpIDV.....	72
Contribution of authors.....	72
Abstract.....	72
Résumé.....	72
Introduction.....	73
Results.....	76
Achnowledgments.....	83
Chapter V: Cricket densoviruses.....	84
Preamble.....	86
Article 5: The <i>Acheta domesticus</i> densovirus, isolated from the European house cricket, has evolved an expression strategy unique among parvoviruses	88
Contribution of authors.....	88
Abstract.....	89
Résumé.....	90
Introduction.....	91
Materials and methods.....	93
Results.....	98
Discussion.....	111
Achnowledgments.....	113
Supplemental materials.....	114
Article 6: A Novel Ambisense Densovirus, <i>Acheta domesticus</i> Mini Ambidensovirus, from Crickets.....	120
Contribution of authors.....	120
Abstract.....	120
Résumé.....	120
Results.....	121
Achnowledgments.....	123
In progress: Novel densoviruses from crickets include a segmented densovirus.....	124
Abstract.....	124

Results.....	125
Discussion.....	129
Chapter VI: Discussion	132
1. The discovery of new iteradensoviruses.....	134
2. Cloning and sequencing of new iteradensoviruses genome.....	135
3. Expression strategy of iteradensoviruses.....	138
4. The unique transcription strategy of AdDNV.....	141
5. Comparison of AdDNV and sAAV capsid structures.....	142
6. The cricket densoviruses and other novel single-stranded DNA viruses of insects	145
7. Horizontal transmission of cricket densovirus.....	146
Conclusions and Perspectives.....	150
References	154
Appendix I: The Structure of Densovirus and Parvovirus.....	174
Article 7: The Structure and Host Entry of an Invertebrate Parvovirus.....	176
Contribution of authors.....	176
Abstract.....	177
Résumé.....	177
Introduction.....	178
Materials and methods.....	182
Results and Discussion	188
Achnowledgments.....	198
Article 8: The structure of Serpine AAV Conserves the Capsid Topology of the Vertebrate AAVs.....	200
Contribution of authors.....	200
Abstract.....	201
Résumé.....	201

Results.....	202
Appendix II: Novel single-stranded DNA viruses	206
Article 9: Organization of the Ambisense Genome of the <i>Helicoverpa armigera</i>	
Densovirus.....	208
Contribution of authors.....	208
Abstract.....	208
Résumé.....	208
Results.....	209
Article 10: <i>Pseudoplusia includens</i> Densovirus Genome Organization and Expression	
Strategy.....	212
Contribution of authors.....	212
Abstract.....	212
Résumé.....	212
Results.....	213
Achnowledgments.....	215
Article 11: A Circo-Like Virus Isolated from <i>Penaeus monodon</i> Shrimps..... 216	
Contribution of authors.....	216
Abstract.....	216
Résumé.....	216
Results.....	217
Achnowledgments.....	218
Appendix III: Publications	220

LIST OF TABLES

Introduction

Table 1.1 New classification of <i>Densovirinae</i> subfamily.....	9
Table 1.2 Five new classified genera of <i>Densovirinae</i> subfamily.....	11

Chapter IV

Table 4.1 Primers for the experiments.....	77
--	----

Chapter V

Table 5.1 PCR primers used in this study.....	103
---	-----

Appendix I

Table a1.1 Structural studies of autonomous parvoviruses.....	180
Table a1.2 X-ray data collection and structure refinement.....	184
Table a1.3 Sequence and structural comparisons of AdDNV capsid protein with other autonomous parvovirus capsid proteins.....	189
Table a1.4 Superposition of the AdDNV jelly roll core (70 residues) on other invertebrate parvovirus capsid proteins.....	189
Table a1.5 Results of a six-dimensional search on fitting of the AdDNV capsid protein structure into the 5.5-Å cryoEM density by using the EMfit program (55)a.....	196
Table a1.6 AAV sequence identities.....	203

LIST OF FIGURES

Introduction

Figure 1.1 Densovirus for biology control.....	6
Figure 1.2 ORFs of ambidensoviruses.....	13
Figure 1.3 Genome organization of iteradensoviruses.....	15
Figure 1.4 Porcine parvovirus transcription map.....	18
Figure 1.5 Details of MIDNV, AaIDNV and BgDNV transcription map.....	23
Figure 1.6 Hairpin structure of different densoviruses.....	26
Figure 1.7 Conserved regions between densoviruses and vertebrate parvoviruses.....	28
Figure 1.8 Parvoviral Rolling hairpin replication (RHR) scheme.....	31
Figure 1.9 Conserved motifs within the viral genome and PLA2 motif.....	34
Figure 1.10 The architecture of parvovirus.....	36
Figure 1.11 Structure comparison of three densoviruses.....	39
Figure 1.12 Adeno-associated virus (AAV) capsid structure.....	41

Chapter II

Figure 2.1 Model for identifying new viruses' genome sequence from SISPA results.....	46
---	----

Chapter III

Figure 3.1 EM of SfDV virus particle and J-shaped telomere sequence of SfDV ITR.....	59
Figure 3.2 EM of PpDV virus particle and J-shaped telomere sequence of PpDV.....	63
Figure 3.3 EM of DppIDV virus particle and J-shaped telomere sequence of DppIDV.....	67

Chapter IV

Figure 4.1 Iteradensoviruses and their transcription.....	75
Figure 4.2 Genome organization and transcription profiles of 5 iteradensoviruses.....	79
Figure 4.3 Expression of non-structural proteins NS1 and NS2 of SfDV.....	82

Chapter V

Figure 5.1 Genome organization of AdDNV.....	100
Figure 5.2 SDS-PAGE, Northern-blotting and promoter activities.....	102
Figure 5.3 Results for mRNA sequencing and mass spectrometry.....	105
Figure 5.4 Conformation of the splicing between ORF-A and ORF-B.....	106
Figure 5.5 Expression of pFastbac VP constructs.....	108
Figure 5.6 Potential VP1 proteins encoded by the VP gene cassette.....	110
Figure S1: Annotated sequence of the AdDNV strain isolated in 1977.....	116
Figure S2: MS results.....	117
Figure S3: Similarities of AdDNV ORF sequences with those of other densoviruses.....	118
Figure 5.7: Genome analysis of AdSDV.....	126
Figure 5.8: Telomeres and genome organization of AdSDV.....	127
Figure 5.9: The new recombinant cricket aphrodisiac iridescent virus.....	129
Figure 5.10: Hypothesis on the origin of AdSDV.....	131

Chapter VI: Discussion

Figure 6.1 Genome comparison of iteradensoviruses.....	137
Figure 6.2 Structure comparison of AdDNV and sAAV.....	144

Appendix I

Figure a1.1 Structure of the AdDNV capsid protein.....	181
Figure a1.2 Heat treatment and PLA2 activity of the virus particles.....	183
Figure a1.3 SDS-PAGE of heat-treated AdDNV particles.....	186
Figure a1.4 Fourier shell correlation (FSC).....	187
Figure a1.5 Structure of AdDNV emptied particles.....	191
Figure a1.6 Glycine-rich regions comparison.....	192
Figure a1.7 Phylogenetic analysis.....	193
Figure a1.8 Structure of the three ordered ssDNA bases in AdDNV particles.....	197
Figure a1.9 Phylogenetic Tree for AAV1-AAV13 and sAAV.....	202
Figure a1.10 Pseudo-atomic model of sAAV	204
Figure a1.11 AAV capsid structures	205

LIST OF ABBREVIATIONS

3D	Three-dimensional
Å	ångström
aa / a.a.	Amino acid
AaeDNV	<i>Aedes aegypti</i> densovirus
AalDNV	<i>Aedes albopictus</i> densovirus
AAV	<i>Adeno-associated</i> virus
AdDNV	<i>Acheta domesticus</i> densovirus
AdMADV	<i>Acheta domesticus</i> mini ambidensovirus
AdSDV	<i>Acheta domesticus</i> segment densovirus
AgDNV	<i>Anopheles gambiae</i> densovirus
AMV	Avian myeloblastosis virus
ATP	Adenosine triphosphate
B19	human parvovirus B19
BgDV1	<i>Blattella germanica</i> densovirus 1
BmDNV-1 (BmDV)	<i>Bombyx mori</i> densovirus
CeDNV (CeDV)	<i>Casphalia extranea</i> densovirus
CpDNV (CpDV)	<i>Culex pipens</i> densovirus
CppDNV	<i>Culex pipiens pallens</i> densovirus
DNVs (DVs)	Densoviruses
DpDNV	<i>Dendrolimus punctatus</i> densovirus

DPE	Downstream promoter element
DppIDV	<i>Danaus plexippus plexippus</i> iteradenvirus
DMSO	Dimethyl sulfoxide
FBS	Foetal bovine serum
FPV	Feline parvovirus
g (µg,mg)	Gram
GFP	Green fluorescence protein
GmDNV	<i>Galleria mellonella</i> densovirus
HeDNV	<i>Haemagogus equinus</i> densovirus
hr	Hour
ICTV	International Committee on Taxonomy of Viruses
IF	Immunofluorescence
IHHNV	Infectious Hypodermal and Hematopoietic Necrosis Virus
Inr	Initiator
ITRs	Inverted Terminal Repeats
JcDNV (JcDV)	<i>Junonia coenia</i> densovirus
KDa	Kilo Dalton
MIDNV (MIDV)	<i>Mythimna loreyi</i> densovirus
MpDNV	<i>Myzus persicae</i> densovirus
MS	Mass spectrometry
MVM	Minute virus of mice

NS	Non-structural protein
nt (nts)	Nucleotide(s)
ORFs	Open reading frames
PBS	Phosphate-buffered saline
PcDV	<i>Planococcus citri</i> densovirus
PfDNV (PfDV)	<i>Periplaneta fuliginosa</i> densovirus
PiDNV	<i>Pseudoplusia includens</i> densovirus
PLA2	Phospholipase A2
PpDNV (PpDV)	<i>Papilio polyxenes</i> densovirus
PPV	Porcine parvovirus
PstDNV (PstDV1)	<i>Penaeus stylirostris</i> penstyldensovirus
RACE	Rapid Amplification of cDNA Ends
RCR	Rolling-circle replication
RCRE	Rolling-circle replication initiation endonuclease
RF	Replicative-form
RHR	Rolling hairpin replication
SF3H	Superfamily 3 helicase
SfDNV (SfDV)	<i>Sibine fusca</i> densovirus
ssDNA	Single-stranded DNA
wt	Wild-type
VP	Viral protein
VRs	Variable regions

Chapter I: Introduction

1 Viruses and biological control

Arthropods are undoubtedly the most widespread and diverse groups of animals, with an estimated 4–6 million species worldwide (Novotny *et al.*, 2002). While only a small percentage of arthropods are classified as pest species, they nevertheless cause major devastation to crops, destroying around 18% of the world's annual crop production (Oerke *et al.*, 2004), contributing to losses of nearly 20% of stored food grains (Bergvinson *et al.*, 2004), and causing around US\$100 billion in damage each year (Carlini *et al.*, 2002). In 2001, a total of US\$7.56 billion was spent in order to protect crops from damage by invertebrate phytophagous pest species (Nicholson, 2007).

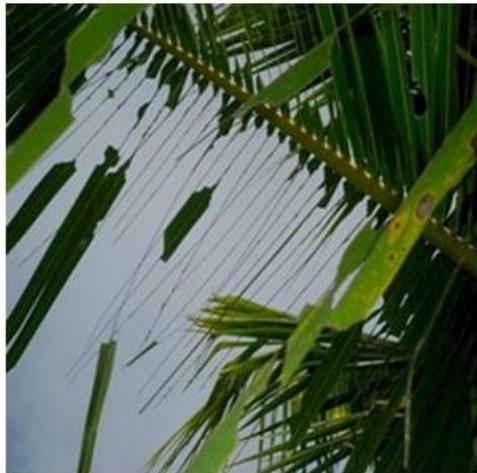
Pest control is a worldwide problem in agricultural and forest ecosystem management. Broad spectrum chemical pesticides have been used abundantly in the containment and eradication of pests of medical, veterinary, agricultural and environmental importance. Though chemical crop protection plays an important role in modern agricultural practices (Fest *et al.*, 1982), it is still viewed as a profit-induced poisoning of the environment. The nondegradable chemical residues, which accumulate to harmful levels, are the root cause of health and environmental hazards and deserve most of the present hostility toward them. Meanwhile, the widespread use of classical agrochemical pesticides, along with their limited number of nervous system targets, has inevitably resulted in widespread resistance among arthropod populations. Moreover, synthetic pesticides often disrupt the balanced insect communities (Kurstak, 1982). This leads to the interest in biological control methods for vectors and plant pests (De Bach, 1964).

Recent advances have highlighted the potential of genetic engineering in the development of novel bio-insecticides which utilize natural organisms, or their products, in the production of transgenic plants or recombinant baculoviruses (Nicholson, 2007). Baculoviruses are arthropod-specific, unable to infect vertebrates or plants (Herniou *et*

al., 2003), and have been applied to protect crops in the agricultural sector since the 1930s (Inceoglu *et al.*, 2001). However, they are limited by their slow ‘time-to-kill’: it may take days to weeks after infection from applying virus to stop feeding, with consequent further damage to the crop (Bonning *et al.*, 1996). The development of recombinant baculoviruses has greatly improved the speed of action via the insertion of foreign gene products. However, the public reticence of the use of genetically modified organisms has contributed to their limited commercial development. Thus, viruses that act rapidly will be of great importance as viable alternatives to chemical pesticides. As an example, the polydnnaviruses, which are introduced into insect larvae through parasitism to temporarily reduce or knock out insect immunity in favour of a parasitoid, are potential candidates for biological control with insect viruses (Lavine *et al.*, 1995).

The densovirus are another group of viruses with great potential for biological control due to their high virulence and infectivity for their natural hosts, although their use as viral pesticides has so far been limited. Field trials to control different species of mosquitos with a Densovirus infecting the mosquito *Aedes aegypti* showed that the virus had a significant efficiency (Buchatsky *et al.*, 1987). The insecticidal potential of *Galleria mellonella* Densovirus on its host, the greater wax moth, has also been studied (Tal *et al.*, 1993). Densovirus were also successfully used for controlling the pests in oil palm fields in Colombia and Ivory Coast infested with *Sibine fusca* and *Caspalia extranea* (Fédière *et al.*, 2002), respectively (Figure 1.1). In this case, application of as few as 10-50 infected larvae per hectare was sufficient to achieve over 90% protection (Belloncik, 1989, Genty *et al.*, 1975). The virus can spread rapidly to adjacent parcels, and the larvae stop feeding quite fast, usually after 1-2 days and start showing paralysis only 8-10 days after infection. Although it does not carry over very well during the wet and dry seasons in West Africa, this successful method of using densovirus for biological control showed great potential.

A



B



Figure 1.1 Densovirus for biology control.

Casphalia extranea densovirus (CeDNV) used for biological control in oil palm fields of Ivory Coast. A) Oil palm leaves destroyed by CeDNV. B) Helicopters were used for spraying the virus.

2 Insect parvoviruses (Densoviruses)

2.1 Genotype and Taxonomic structural history of Densoviruses

The first densovirus was discovered in 1964 when Meynadier and his colleagues observed a high mortality within an insect rearing facility of larvae of the greater wax moth *G. mellonella* in their research unit at St-Christol (France). They observed an outbreak caused by a new virus that was identified as a densonucleosis virus according to its cytopathological symptoms (Meynadier *et al.*, 1964). A second densovirus,

BmDNV-1, was discovered just after the GmDNV isolation in sericultural farms in Ina City, Japan (Kawase, 1985, Kawase *et al.*, 1976). Since then, an increasing number of densoviruses have been reported as a group of pathogens that are highly pathogenic and fatal to their hosts.

Densoviruses (DNVs) are parvoviruses of invertebrates, infecting at least insects, shrimps, and urchins, and form a separate subfamily (*Densovirinae*) within the *Parvoviridae* family (Cotmore *et al.*, 2014a, Gudenkauf *et al.*, 2014, Tijssen *et al.*, 2011). Nevertheless, DNVs share many properties with the *Parvovirinae* subfamily of vertebrates such as a single-stranded (ss), linear DNA genome of about 4-6 kb, terminal hairpins in the genomes which are for some viruses in the form of inverted terminal repeats (ITRs) and similar icosahedral capsid structures consisting of 60 proteins some of which have N-terminal extensions (Chapman *et al.*, 2006, Cotmore *et al.*, 2006a, Tijssen *et al.*, 2006a).

Together with the *Parvovirinae* subfamily of vertebrates, with which they share little sequence identity, the *Densovirinae* make up the family of *Parvoviridae* (Berns *et al.*, 2000). Thus far, a limited number of around thirty densovirus were divided into three genera, *Densovirus*, *Iteravirus* and *Brevidensovirus* (Anonyme, 2000). Due to the growing knowledge about these densoviruses, the taxonomy of *Densovirinae* subfamily was last modified in 2004, prior to publication of the 8th International Committee on Taxonomy of Viruses (ICTV) Report (Tattersall *et al.*, 2005). According to this reclassification, the genus *Densovirus* was divided into three subgroups (subgroup A, B and C) and a single new genus, the *Penfudensovirus* was introduced.

In 2014, a set of proposals to update the taxonomy of the family *Parvoviridae* has been submitted by the ICTV *Parvoviridae* Study Group (Cotmore *et al.*, 2014a), and have been approved by ICTV. The new taxonomy followed a novel viral definition rule: "In order for an agent to be classified in the family *Parvoviridae*, it must be judged to be an authentic parvovirus on the basis of having been isolated and sequenced or,

failing this, on the basis of having been sequenced in tissues, secretions, or excretions of unambiguous host origin, supported by evidence of its distribution in multiple individual hosts in a pattern that is compatible with dissemination by infection. The sequence must be in one piece, contain all the non-structural (NS) and viral particle (VP) coding regions, and meet the size constraints and motif patterns characteristic of the family" (Cotmore *et al.*, 2014a). In the subfamily *Densovirinae*, proposed changes include the introduction of two new genera for shrimp viruses and the expansion of the existing genus names *Iteravirus* and *Densovirus* to *Iteradensovirus* and *Ambidensovirus*, respectively. In both genera, species identity levels have been lowered, numbered, binomial species names adopted, and new species introduced (Cotmore *et al.*, 2014a). For viruses from the subfamily *Densovirinae*, viral names have typically been assembled from binomial host names plus the word "densovirus", for example, *Junonia coenia* densovirus, originally abbreviated to JcDNV (where the capitalized "N" harks back to a time when these viruses were called "densonucleosis viruses"). The Study Group suggested eliminating the vestigial N from all abbreviations (Table 1.1). The new taxonomy of subfamily *Densovirinae* includes genera *Ambidensovirus*, *Iteradensovirus*, *Brevidensovirus* and two new genera for shrimp viruses: *Hepandensovirus* and *Penstyldensovirus*.

Table 1.1 New classification of *Densovirinae* subfamily.*

Genus	Species	Virus or virus variants	Abbreviation	Accession #
<i>Ambidensovirus</i>	<i>Blattodean ambidensovirus 1</i>	Periplaneta fuliginosa densovirus	PfDV	AF192260
	<i>Blattodean ambidensovirus 2</i>	Blattella germanica densovirus 1	BgDV1	AY189948
	<i>Dipteran ambidensovirus 1</i>	Culex pipens densovirus	CpDV	FJ810126
	<i>Hemipteran ambidensovirus 1</i>	Planococcus citri densovirus	PcDV	AY032882
	<i>Lepidopteran ambidensovirus 1</i>	Diatraea saccharalis densovirus	DsDV	AF036333
		Galleria mellonella densovirus	GmDV	L32896
		Helicoverpa armigera densovirus	HaDV1	JQ894784
		Junonia coenia densovirus	JcDV	S47266
		Mythimna loreyi densovirus	MIDV	AY461507
		Pseudoplusia includens densovirus	PiDV	JX645046
<i>Brevidensovirus</i>	<i>Orthopteran ambidensovirus 1</i>	Acheta domesticus densovirus	AdDV	HQ827781
	<i>Dipteran brevidensovirus 1</i>	Aedes aegypti densovirus 1	AaeDV1	M37899
		Aedes albopictus densovirus 1	AalDV1	AY095351
		Culex pipiens pallens densovirus	CppDV	EF579756
		Anopheles gambiae densovirus	AgDV	EU233812
		Aedes aegypti densovirus 2	AaeDV2	FJ360744
	<i>Dipteran brevidensovirus 2</i>	Aedes albopictus densovirus 2	AalDV2	X74945
		Aedes albopictus densovirus 3	AalDV3	AY310877
		Haemagogus equinus densovirus	HeDV	AY605055
		Penaeus monodon hepandensovirus 1	PmoHDV1	DQ002873
<i>Hepandensovirus^a</i>	<i>Decapod hepandensovirus 1</i>	Penaeus chinensis hepandensovirus	PchDV	AY008257
		Penaeus monodon hepandensovirus 2	PmoHDV2	EU247528
		Penaeus monodon hepandensovirus 3	PmoHDV3	EU588991
		Penaeus merguiensis hepandensovirus	PmeDV	DQ458781
		Penaeus monodon hepandensovirus 4	PmoHDV4	FJ410797
		Fenneropenaeus chinensis hepandensovirus	FchDV	JN082231
	<i>Lepidopteran iteradensovirus 1</i>	Bombyx mori densovirus	BmDV	AY033435
	<i>Lepidopteran iteradensovirus 2</i>	Casphalia extranea densovirus	CeDV	AF375296
		Sibine fusca densovirus	SfDV	JX020762
	<i>Lepidopteran iteradensovirus 3</i>	Dendrolimus punctatus densovirus	DpDV	AY665654
<i>Iteradensovirus</i>	<i>Lepidopteran iteradensovirus 4</i>	Papilio polyxenes densovirus	PpDV	JX110122
	<i>Lepidopteran iteradensovirus 5</i>	Helicoverpa armigera densovirus	HaDV2	HQ613271
	<i>Decapod penstyldensovirus 1</i>	Penaeus stylostris penstyldensovirus 1	PstDV1	AF273215
		Penaeus monodon penstyldensovirus 1	PmoPDV1	GQ411199
		Penaeus monodon penstyldensovirus 2	PmoPDV2	AY124937
		Penaeus stylostris penstyldensovirus 2	PstDV2	GQ475529

The type species for each genus is indicated in bold type

^a Indicates genus of viruses formerly known as hepatopancreatic parvovirus [HPV] of shrimp

^b Indicates genus of viruses formerly known as infectious hypodermal and hematopoietic necrosis virus (IHHNV) of shrimp

* According to the new classification of *Densovirinae* subfamily, the abbreviation names for viruses have also been changed. i.e. GmDNV->GmDV, MIDNV->MIDV, CeDNV->CeDV. In this thesis, we will use the new abbreviations in the research results obtained since 2014. For the viruses reported before the newly released classification of 2014, we will still keep the original abbreviation.

2.2 Pathology caused by densovirus

The densovirus are responsible for fatal diseases of their host. Mortality due to densovirus infections takes effect between two to twenty days after infection depending on the inoculating virus concentration and the insect larval stage (Meynadier *et al.*, 1964, Shimizu, 1975). Generally the first symptoms are anorexia and lethargy followed by flaccidity and the inhibition of moulting and metamorphosis. During infection, larvae become whitish and progressively paralyzed, followed by a slow melanisation (Meynadier *et al.*, 1964, Vago *et al.*, 1966). Insect larvae usually develop anorexia a few days after infection and stop feeding (Kawase *et al.*, 1990). Densovirus of *Sibine fusca* and *Casphalia extranea* produce tumor-like lesions in the intestines of their hosts. The midgut epithelial cells of diseased larvae undergo intensive proliferation and the progressive thickening and opacity of the gut wall screens off the intestinal content (Fédière, 2000). In the case of *C. extranea*, the larval color changes from green to yellowish brown and the transparent gut becomes opaque (Fédière, 2000). With BmDNV-1 infected silkworm larvae, the alimental canal of the diseased larvae is pale yellow with little internal content and the larvae usually die after seven days showing body flaccidity (Shimizu, 1975).

2.3 Genome organization of densovirus

Densovirus package a monopartite linear single-stranded DNA molecule (ssDNA) of 4-6 kb within their virion of either a negative or a positive polarity. The proportion of encapsidated negative and positive strands of DNA varies according to the virus genus. For members of the *Iteradenovirus* genus, both strands are encapsidated in an equimolar proportion sharing that characteristic with vertebrate dependoviruses (AAVs) (Berns *et al.*, 1972). In contrast, in *Breviadenovirus*, AaeDNV and AalDNV, the negative strand is preferentially encapsidated (90%) similarly to vertebrate parvovirus such as MVM and PPV (Muzyczka *et al.*, 2001, Tijssen, 1995). Cloning of the complete genome of several densovirus and determination of their sequence has provided a prediction of the genomic organization of this group of viruses.

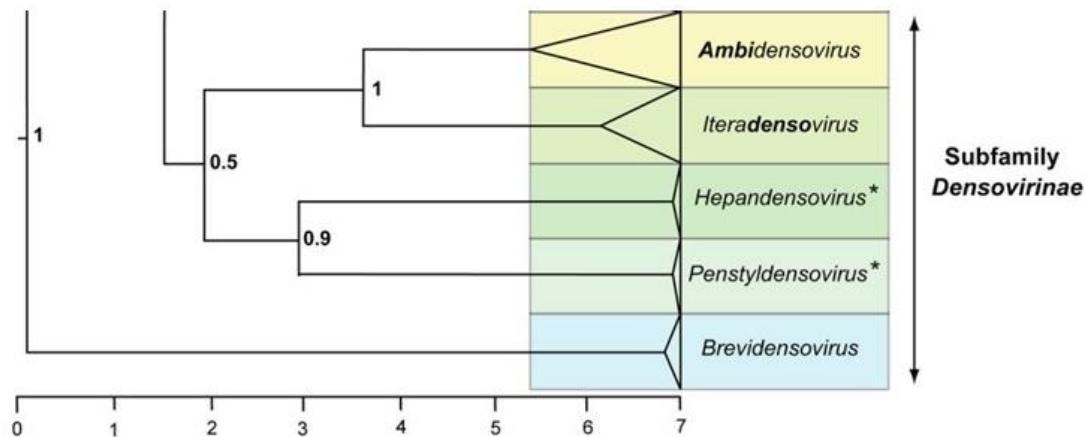
The emerging number of new densoviruses led to a reclassification of their taxonomy (Cotmore *et al.*, 2014a) (Table 1.2).

Table 1.2 Five new classified genera of *Densovirinae* subfamily.

A

Current taxonomy			Proposed taxonomy		
Genus	# Species	# viruses or strains	Genus	# Species	# viruses or variants
Subfamily <i>Densovirinae</i> - arthropod hosts					
<i>Densovirus</i>	2	2	<i>Ambidensovirus</i>	6	11
<i>Brevidensovirus</i>	2	2	<i>Brevidensovirus</i>	2	8
-	-	-	<i>Hepandensovirus^a</i>	1	7
<i>Iteravirus</i>	1	1	<i>Iteradensovirus</i>	5	6
-	-	-	<i>Penstyldensovirus^a</i>	1	4
<i>Pefudenvirus^b</i>	1	1	-	-	-

B



A) Summary of changes between the previous and new taxonomy. B) Phylogenetic tree showing genera in the subfamily *Densovirinae*. Phylogenetic analysis is based on the amino acid sequence of the viral replication initiator protein, NS1, which contains a conserved AAA+ helicase domain corresponding to the Parvo_NS1 Pfam domain: http://pfam.sanger.ac.uk/family/Parvo_NS1. The size of the color block for each genus indicates the relative number of species currently recognized, as an indicator of its diversity (Cotmore *et al.*, 2014a).

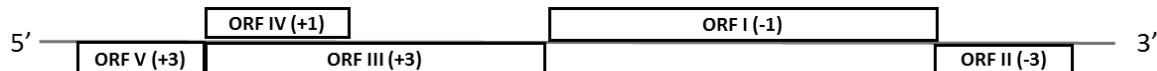
2.3.1 Genome organization of the *Ambidensovirus* genus

This genus includes all densoviruses having an ambisense genome organization, with the VP (structural protein) and NS (non-structural protein or Rep protein) on the 5' halves of the complementary strands, and with inverted terminal repeats (ITRs) (Tijssen *et al.*, 2006a). Members were sub-divided into six host species. *Blattodean ambidensovirus* 1 and *Blattodean ambidensovirus* 2 are two cockroach host species include PfDV and BgDV1 (Mukha *et al.*, 2006), respectively. *Dipteran ambidensovirus* 1 includes the mosquito ambidensovirus CpDV. PcDV (Thao *et al.*, 2001) is a densovirus infecting mealybugs and belongs to *Hemipteran ambidensovirus* 1. The Lepidopteran ambidensoviruses contain the classical ambisense densoviruses usually isolated from lepidopteran insects such as JcDV. The densovirus isolated from cricket host *Acheta domesticus* first belonging to a separate genus *Pefudensovirus* was reclassified as the sixth host species group of *Ambidensovirus* genus. Figure 1.2 shows the comparison of the open reading frames (ORFs) for each of the classical members.

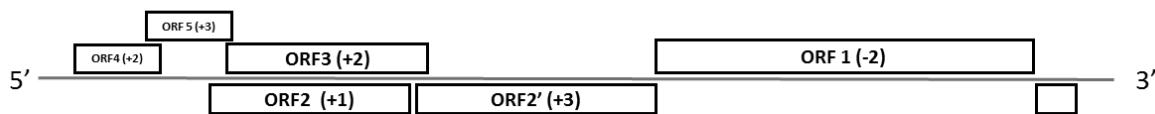
Blattodean ambidensovirus 1 (*Periplaneta fuliginosa* Densovirus) (5,454 nt)



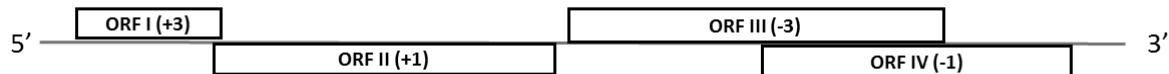
Blattodean ambidensovirus 2 (*Blattella germanica* densovirus 1) (5,335 nt)



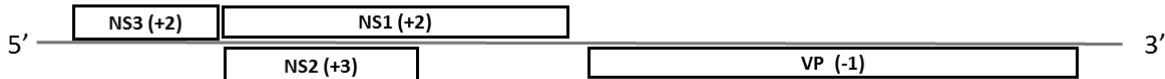
Dipteran ambidensovirus 1 (*Culex pipens* Densovirus) (5,759 nt)



Hemipteran ambidensovirus 1 (*Planococcus citri* Densovirus) (5,380 nt)



Lepidopteran ambidensovirus 1 (*Junonia coenia* densovirus) (6,032 nt)



Orthopteran ambidensovirus 1 (*Acheta domesticus* Densovirus) (5,425 nt)

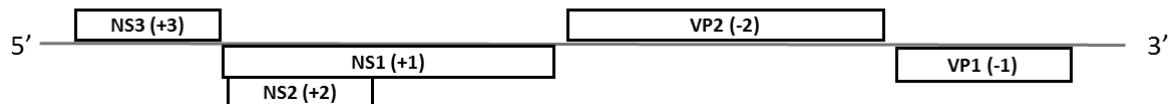


Figure 1.2 ORFs of ambidensoviruses.

The large open reading frames (ORFs) for the member from different group of *Ambidensovirus* genus. The graphical analysis tool ‘ORF Finder’ <http://www.ncbi.nlm.nih.gov/gorf/orfig.cgi> has been used for analyze virus genome sequence. The ‘+’ or ‘-’ with number indicate different frames of complement strands.

Both of the complementary DNA strands of ambisense densoviruses have the capacity to code for proteins through a limited number of large ORFs distributed within the 5'-half of the complementary strands. One strand codes for structural proteins, usually 4 (VP1-VP4) (Tijssen et al., 1976), whereas the complementary strand codes for the three non-structural proteins NS1-NS3. As exemplified by GmDNV, which is the first densovirus isolated from insects (Meynadier et al., 1964), the members from *Lepidopteran ambidensovirus 1* exhibit an identical organization of their genome. The genome of GmDNV contains a long inverted terminal repeat (ITR) at both ends of about 550 nts. The palindromic sequence of 136 nts folds back to create a Y-shaped hairpin with flip and flop orientations (Tijssen et al., 2003). Minus and plus strands package in separate virions in a 1/1 ratio. Three ORFs coding for three NS proteins are present on one strand, the left-most ORF (near the 5' extremity) being in frame with, and separated from the largest NS ORF only by a TAA stop codon. The largest ORF codes for NS1 (Rep protein) and contains the NTPase motif that is typically found in NS1 of all parvoviruses. A single large ORF codes for the four structural polypeptides on the complementary strand. The VP gene containing phospholipase A2 (PLA2) motif is within the N-terminus of the VP protein. The genome of AdDNV is different in that (i) three (two of which overlap) ORFs are present on one strand, each likely coding for a single NS protein, and (ii) the VP gene coding for the four structural polypeptides on the complementary strand is split in two and, successively with two reading frames (Liu et al., 2011) (Figure 1.2).

Other Lepidopteran densoviruses such as *Diatraea saccharalis* Densovirus DsDNV, *Pseudoplusia includens* Densovirus PiDNV and *Junonia coenia* Densovirus JcDNV share the same characteristics and high sequence identity with GmDNV. However, other ambidensoviruses isolated from orders other than Lepidoptera (Table 1.1), such as PfDNV and CpDNV are not homogenous in genome organization. These viruses have a shorter genome of around 5.5 kb and smaller structural protein molar masses as well as relative amounts that are distinct from the classical members (Figure 1.2). They also have shorter ITRs that are different in both size and structure.

2.3.2 Genome organization of *Iteradensovirus* genus

All members in the *Iteradensovirus* genus possess a monosense genomic organization, i.e. their sequences coding for NS and VP proteins are located in the left and the right halves of the same strand. Until a few years ago, only three iteradensoviruses were known: BmDNV-1 from *Bombyx mori* (Nakagaki *et al.*, 1980), CeDNV from *Casphalia extranea* (Fédière *et al.*, 2002) and DpDNV from *Dendrolimus punctatus* (Wang *et al.*, 2005). Their genomes are about 5 kb in length with 70-85% sequence identity, particularly in their typical J-shaped ITRs. Recently, a second densovirus HaDNV-1 was reported in *Helicoverpa armigera* which was clustered into this genus by phylogenetic analysis but complete ITRs sequences were not been obtained (Xu *et al.*, 2012). All the iteradensovirus genomes contain three large open reading frames, two overlap in the left-hand half of the genome and encoding non-structural proteins (NS1 and NS2) and one in the right-hand half encoding capsid proteins (VP1-VP4) (Figure 1.3). NS1 protein contains replicator protein and helicase/ATPase motifs. The GPG and HD regions in VP1 capsid protein is conserved among most parvoviruses and contain the Ca²⁺ binding motif and phospholipase A2 (pLA2) motif.

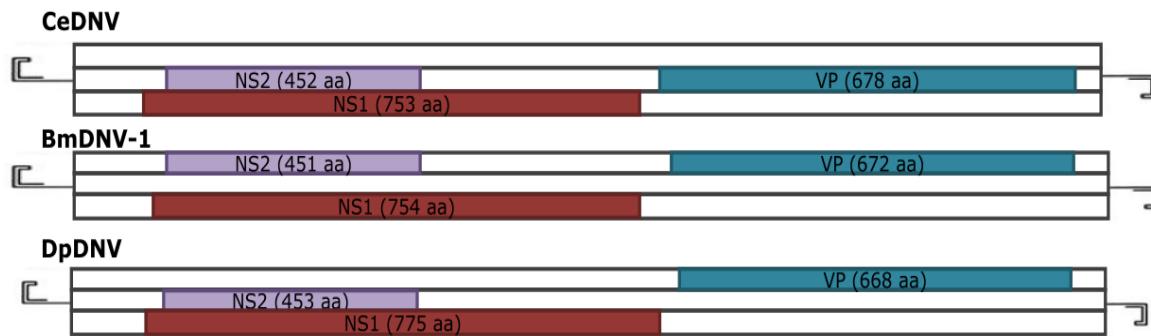


Figure 1.3 Genome organizations of iteradensoviruses.

Comparison of genome organizations for *Casphalia extranea* densovirus CeDNV, *Bombyx mori* densovirus BmDNV-1 and *Dendrolimus punctatus* densovirus DpDNV.

2.3.3 Genome organization of the *Brevidensovirus* genus

Most members in this group are isolated from persistently or chronically infected mosquito cell lines: *Aedes aegypti* Densovirus AaeDNV (Afanasiev et al., 1991), *Aedes albopictus* Densovirus AalDNV (Boublik et al., 1994), *Anopheles gambiae* Densovirus AgDNV (Ren et al., 2008), *Culex pipiens pallens* densovirus CppDNV (Zhai et al., 2008), and *Haemagogus equines* Densovirus HeDNV (Paterson et al., 2005). Their hosts are known as the epidemiologically essential vectors for several dangerous human pathogens such as yellow fever virus, dengue fever virus, malaria, chikungunya virus, and West Nile virus. Therefore, mosquito brevidensoviruses are potential biological insecticides to control these paratransgenic pathogens. Members in *Brevidensovirus* genus contain the smallest genome (about 4 kb) in the *Densovirinae* subfamily. The genomic organization is also monosense, but this genus has some distinct characteristics from other genera. Their genomes do not have ITRs but contain distinct T-shaped hairpin telomeres. They have two large ORFs for NS genes instead of three, as in the case of the *Ambidensovirus* genus. In addation, there is only one large ORF for the VP genes, but it does not contain the PLA2 motif.

2.3.4 Genome organization of the *Hepandensovirus* and *Penstyldensovirus* genus

All densoviruses isolated from shrimp were reclassified into two new genera. *Hepandensovirus* is the genus of viruses formerly known as hepatopancreatic parvovirus [HPV] of shrimp, and *Penstyldensovirus* is a siglum for *Penaeus stylostris* densovirus which formerly known as infectious hypodermal and hematopoietic necrosis virus (IHHNV) of shrimp. The genome of the classical shrimp densovirus (PstDNV = IHHNV) also contains three large ORFs as for the brevidensoviruses. The 5' ORF sequence shares some identity with brevidensovirus NS1 with a rolling-circle replication motif at the N-terminus and an NTP-binding/helicase/NTPase motif at the C-terminus (Pham et al. unpublished data). The upstream, mid, and downstream.

ORFs of the PstDNV sequence (3873 nts, ~98% complete) have potential coding capacities of 666 amino acids (aa), 363 aa, and 329 aa, respectively.

2.4 Transcription strategy of densoviruses

Several models for vertebrate parvovirus transcription profiles have been determined at a detailed level. In each case, the parvoviruses have adapted a complex pattern of alternative splicing and polyadenylation to maximize the information from their small genomes. These complex patterns of expression generally result in a set of mRNAs that express two to three capsid proteins from overlapping ORFs, and one large and two to three small non-structural proteins. In addition, adeno-associated virus (AAV) gene expression has evolved to be dependent on helper-virus function (Qiu *et al.*, 2006). The core promoter element for gene expression usually contains a TATAA consensus sequence and an upstream SP1 binding site with a GC rich motif (Lorson *et al.*, 1996). The activity of this promoter is NS1-dependent. Figure 1.4 illustrates a summary of the porcine parvovirus (PPV) transcription profile. The PPV genome contains two open reading frames (ORFs). The left ORF allows the production of NS proteins from the P4 promoter. The single primary transcript (gray, top) contains two splice sites. The splice site of the small intron is located downstream of the stop codon of the NS protein, and therefore does not change their sequence. The production of the NS1 protein is from non-spliced ORF in the mRNA. Splicing within this ORF allows the production of the NS2 and the sequence after the splice is in a different reading frame relative to NS1. PPV uses a single polyadenylation signal (AAAn) located at the end of the coding regions providing mRNA considerably longer than the coding region for the transcripts of the NS proteins. The structural proteins are produced from a single primary transcript from the right ORF and are transcribed from the p40 promoter. There are two donor sites and one acceptor site for splicing. The first splice donor site removes the start codon of the protein VP1, resulting in the production of the VP2 protein. The second splicing donor site retains the initiation codon for VP1. The VP1 and VP2 proteins come from the same reading frame.

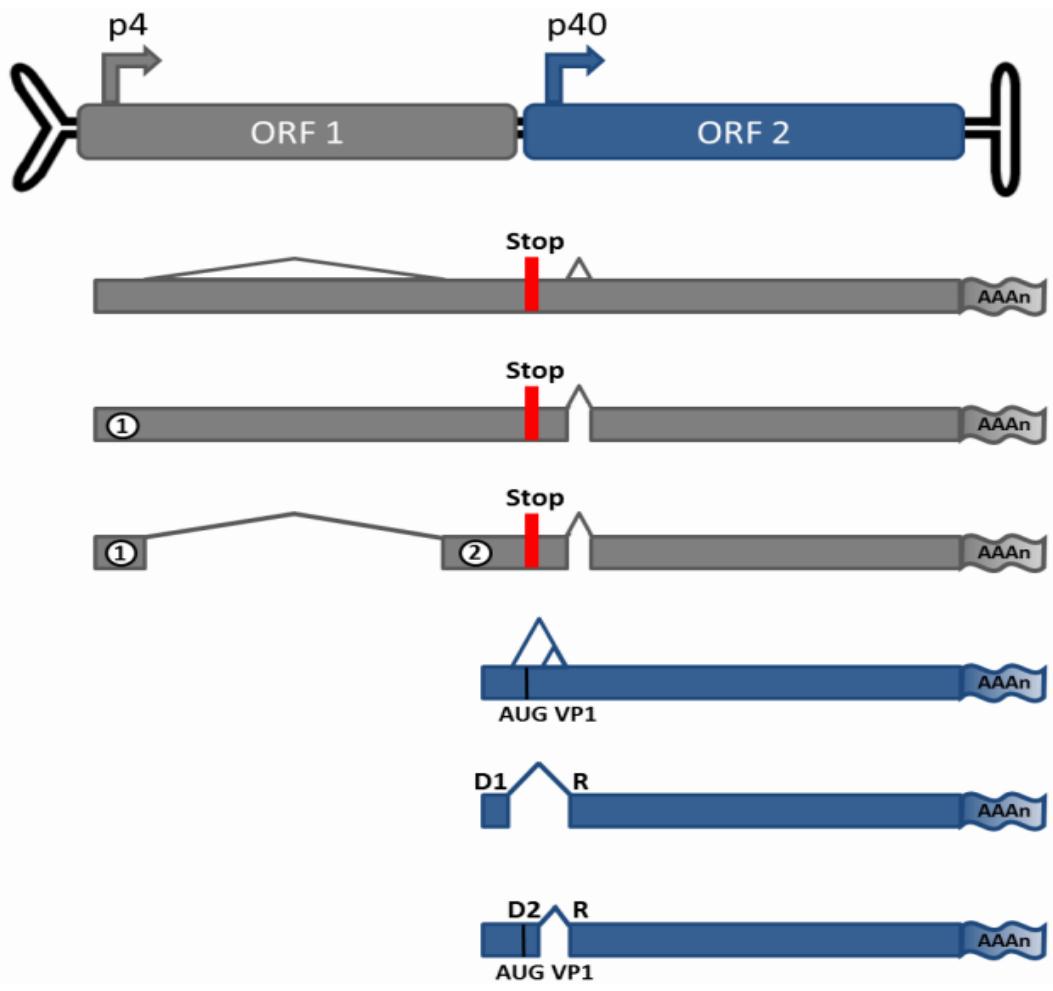


Figure 1.4 Porcine parvovirus transcription map.
D1 and D2 represent the first and second donor sites of VP transcripts, respectively. R represents the receptor sites for splicing (Bergeron *et al.*, 1993).

The expression strategies of densoviruses vary significantly. Densoviruses have been accorded little attention for the gene expression of their genes compared to well-known strategies of the vertebrate parvoviruses. As for most parvoviruses in the *Parvovirinae* subfamily, all densoviral genomes contain at least two distinct promoter sets for non-structural (NS) and structural (VP) genes. Densoviruses also employ alternative splicing, leaky scanning, and polyadenylation mechanisms for producing different proteins from their limited genomes.

Insect gene expression requires transcription factors that are often different from those of vertebrate genes. For instance, insect cells lack Sp1 that recognizes GC-rich domains (Courey *et al.*, 1988, Santoro *et al.*, 1988). Although the TATA box is the most common core promoter element for insect gene expression, almost one half of studied *Drosophila* promoters can be classified as TATA-less promoters (Arkhipova, 1995). A conserved downstream core promoter element (DPE) has been recognized within TATA-less promoters in *Drosophila*, which is required for the sequence-specific binding of the transcription factor, TFIID. The DPE is located from +28 to +32 relative to the +1 transcription start site. Moreover, in *Drosophila*, the DPE appears to be about as common as the TATA box (Kadonaga, 2002). Arthropod initiator elements for transcription starts have a proposed consensus of ATCAG/TTC/T (often CAGT) which allows RNA polymerase II to function even in the absence of a clear TATA-box. The DPE, with the consensus A/G/T-C/G-A/T-C/T-A/C/G-C/T acts in conjunction with the initiator sequence (Inr) (CAGT), to provide a binding site for TFIID in the absence of a TATA box.

To date, the transcription strategies of ambidensoviruses are the most were described. The detailed RNA processing profiles for GmDNV, MIDNV, JcDNV and BgDNV have been reported (Fédière *et al.*, 2004, Kapelinskaya *et al.*, 2011, Tijssen *et al.*, 2003, Wang *et al.*, 2013). Their ambisense nature dictates that transcription occurs on both strands. Indeed, VP and NS genes are transcribed independently, each from one species of un-spliced mRNA, from promoters located within the ITRs (Bergoin *et*

al., 2000). The TATA-box motifs are within the terminal repeats and the CAGT-box in the unique sequence. Figure 1.5A shows the detailed transcription map for MIDNV (Fédière *et al.*, 2004). The NS and VP genes are located on the 5'-halves of the complementary strands and their transcripts have an overlap of 57 nucleotides in the middle of the genome. Two sizes of NS transcripts were detected that start 27 nts downstream of the 5' ITRs. The NS cassette consists of three genes with NS1 and the overlapping NS2 downstream of NS3. Only one promoter, with the TATA-box located within the upstream ITR, drives the expression of the NS transcripts. The NS3 gene was spliced out from a fraction of the NS transcripts to allow leaky scanning translation of the downstream bicistronic NS1 and NS2 genes. The splicing donor site is immediately upstream of the initiation codon of NS3 and the acceptor site immediately upstream of NS1, so that both spliced and unspliced transcripts have an identical leader sequence except for the last two nts (GT) preceding NS3 (Fédière *et al.*, 2004). One large ORF on the right-hand half of the complementary strand codes for approximately four VP proteins. No splicing has been detected within the VP transcripts, suggesting the four VPs were similarly generated by leaky scanning translation of an unspliced mRNA. The 5'-untranslated region of the VP transcript is only seven nucleotides long. The sequence context around the ATG start codons also seems to determine how well that ATG is recognized (Fédière *et al.*, 2004). Thus, the untranslated 5'-terminal sequence of the mRNA coding for the structural polypeptides is very short, so that only a small percentage of the ribosomes initiate translation at the first AUG codon. As a result, VP1 is poorly expressed (Fédière *et al.*, 2004). In contrast, the high level of expression of VP4 is due to the favorable environment of its two initiator codons. Unlike the situation with vertebrate parvoviruses whose mRNAs co-terminate at the 3' extremity of the genome, MIDNV possesses two polyadenylation (poly(A)) signals (AATAAA) in the middle of its genome. The first two nts, AA, in the poly(A) sequence of the NS transcript overlap with the NS stop codon. Therefore, the VP and NS transcripts have 56 nts overlapping sequence. The termination of both the VP and NS transcripts occurs very close to the end of their coding sequences, so that only the 5'-halves of the complementary strands are transcribed. This probably prevents the meeting of the polymerase complexes when

the genes are only partially transcribed (Bergoin *et al.*, 2000). GmDNV and JcDNV have similar genome organization to MIDNV and hence, their expression strategies have only a few differences. In GmDNV and JcDNV, the two sizes of NS transcripts start at 23 and 32 nts downstream of the left ITRs, respectively (Tijssen *et al.*, 2003, Wang *et al.*, 2013). In JcDNV, a single mRNA for expressing VP proteins is transcribed from the P9 promoter. While the P93 is the unique putative promoter on the strand bearing NS genes, a single transcription start for the two NS mRNAs was located 32 nt downstream of the P93 TATA box but 83 nt and 86 nt upstream of NS3 and NS1, respectively. With their TATA box just at the limit of the ITR and their regulatory sequences within the ITR, the P9 and P93 appear to be both structurally and functionally very similar (Wang *et al.*, 2013). (Note: the authors used the original orientation of the JcDNV genome: in the new sequence P9 becomes P91 and P93 becomes P9 (Pham *et al.*, 2013d)).

The *Blattella germanica* densovirus (BgDNV) is an autonomous parvovirus that infects the German cockroach. This virus belongs to the *Ambidensovirus* genus and possesses a 5,335-nt-long ambisense genome that contains five basic ORFs. Two of them (ORF1 and ORF2) on one strand code for VP, and three (ORF3 to ORF5) on another strand code for NS proteins (Kapelinskaya *et al.*, 2008). Moreover, one additional ORF, ORF6, was found to be located on the same strand as ORF3 and ORF5 but within the VP-coding region. The reported expression strategy of BgDNV was unique among the other ambisense densoviruses (Kapelinskaya *et al.*, 2008, Kapelinskaya *et al.*, 2011). BgDNV possesses three mRNAs transcribed from the single P2 promoter for NS proteins, with spliced and unspliced transcript variants. The unspliced variant encodes NS3 which is similar to the classical ambidensoviruses (GmDNV, MIDNV), while the first splice product encodes NS1 and NS2. The second splice product encodes the C-proximal half of NS1. Long NS transcripts have approximately 1,600 nt of overlap with mRNAs for VP proteins. This overlap may lead to the formation of a relatively long region of double-stranded RNA (dsRNA) during the virus life cycle, which in turn can induce a cellular antiviral response by means of RNA interference (Kapelinskaya *et al.*, 2011). BgDNV possesses three VP transcripts which

are transcribed from a single promoter P1, one of which (VP) is unspliced, while the other two (VPspl1 and VPspl2) are generated by alternative splicing. The unspliced VP transcript contains both ORF1 and ORF2, while in VPspl1, ORF1 and ORF2 are joined in frame (Figure 1.5 C). Like the majority of densoviruses, BgDNV possesses only one termination site for mRNAs for both capsid and NS proteins.

The brevidensovirus with monosense genomic organization contain three large ORFs coding for NS and VP gene. Compared to other densoviruses, the transcription pattern of brevidensovirus has not been well characterized. It was reported that PstDNV generates five transcripts: two spliced for NS1, two for NS2 and one for VP under the control of three promoters P2, P11, and P61 (Dhar *et al.*, 2010). Those results have been corrected recently (Pham *et al.* unpublished data). Figure 1.5 B gives a detailed transcription map of AalDNV as an example (Pham *et al.*, 2013a). AalDNV used one promoter region with closely overlapping elements to start transcription of NS1 and NS2 at positions that are just 14 nt apart at either side of ATG_{NS1}. All NS mRNAs co-terminated with VP mRNA. No clear TATA-like motif sequences were found upstream of the initiator sequence CAGT of the VP of AalDNV and AaeDNV, suggesting that these promoters were under the control of DPE (Pham *et al.*, 2013a). This regulatory circuit is likely to be one means by which insect virus networks can transmit transcriptional signals, such as those from DPE-specific and TATA-specific enhancers, via distinct pathways (Hsu *et al.*, 2008), to regulate NS and VP expression.

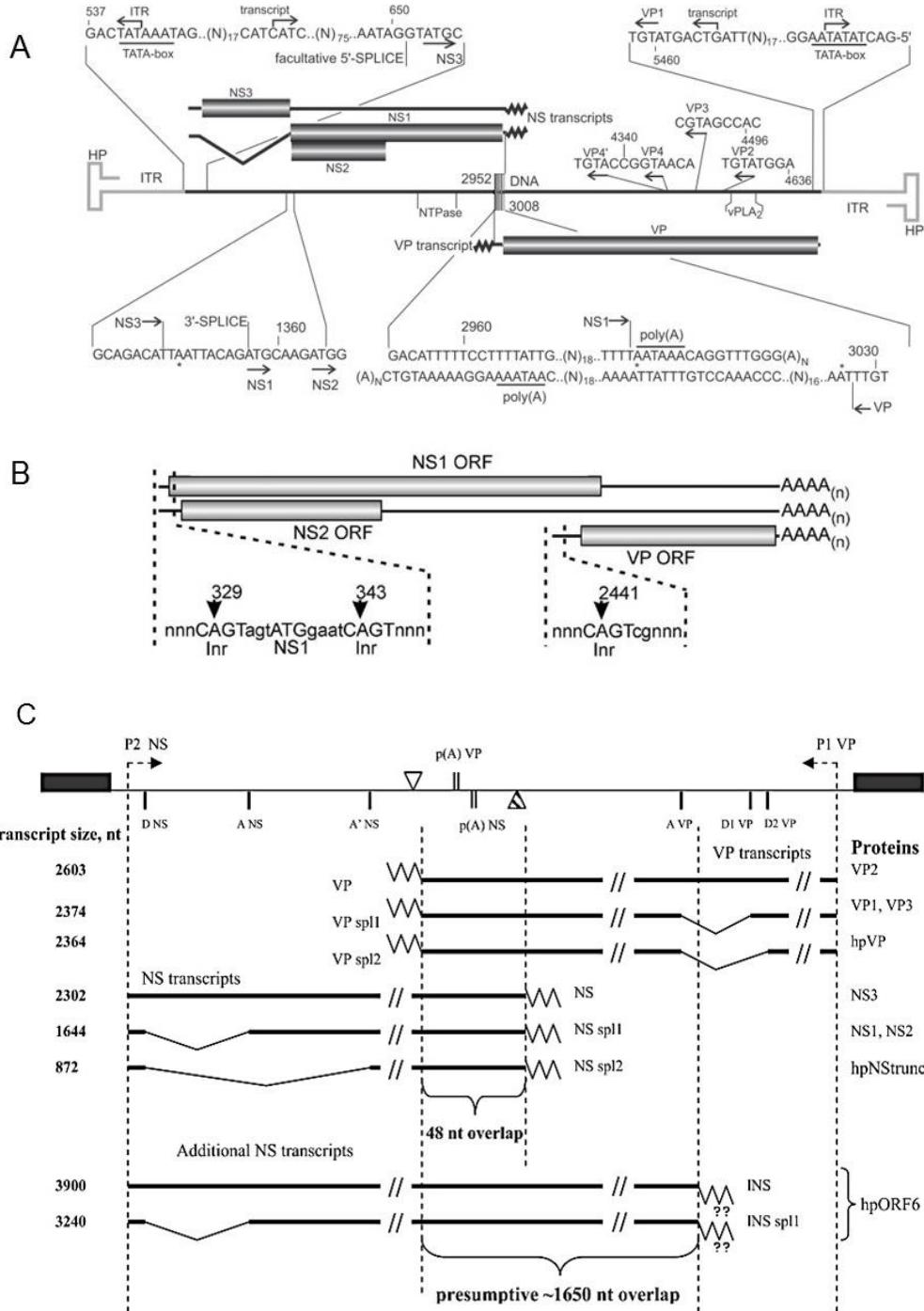


Figure 1.5 Details of MIDNV, AaIDNV and BgDNV transcription map.

A) MIDNV transcription map: two putative promoter regions within the left and right ITRs control the expression of NS and VP transcripts. The 5'-splicing donor and 3'-splice acceptor sites are upstream NS3 and NS1, respectively. The two poly(A) signals are in the middle of the MIDNV genome. The initiation and termination of NS and VP transcripts are also indicated (Fédière *et al.*, 2004). **B) AaIDNV transcription map:** two putative Inr of NS1 and NS2 and one Inr of VP gene. NS and VP gene expression share the same polyadenylation signal downstream of VP (Pham *et al.*, 2013a). **C) Summary of BgDNV genome transcription** (Kapelinskaya *et al.*, 2011).

2.5 Terminal structures of Densovirus genomes

Priming a DNA molecule during the replication process is a key step. DNA viruses have evolved a number of unusual mechanisms to prime the replication to their genome such as RNA priming, DNA priming and even protein priming (Flint *et al.*, 2009). Parvoviruses replicate their short (~ 5 kb) linear single-stranded DNA genomes using a unidirectional strand-displacement strategy called “rolling hairpin” replication, which is an evolutionary adaptation of rolling circle synthesis (Chapman *et al.*, 2006). This adaptation is mediated by small imperfect palindromes positioned at each end of the viral genome, or so-called terminal hairpins, on their 3' and 5' extremities of both complementary strands. This terminal structural contain most of the cis-acting information required for viral DNA replication and packaging (Li *et al.*, 2013). Despite these central roles in replication, the sizes, sequences, and predicted structures of the hairpins can vary remarkably between different genera of the *Parvoviridae* and even between the two ends of a single virus, suggesting they have been adapted to fulfill additional roles in the viral life cycle.

In densoviruses, the presence of these terminal palindromic structures has been proven by both direct observation using electron microscopy as well as by direct sequencing for several viruses and prediction of their secondary structures (Dumas *et al.*, 1992, Fédière *et al.*, 2002, Li *et al.*, 2001, Tijssen *et al.*, 2006a). Those palindromic sequences at the termini can fold into Y-shaped, J-shaped, T-shaped or I-shaped hairpins (Figure 1.6) and serve as primers for viral replication at the 3'-extremity. In addition to their role in viral DNA replication, they allow the rescue of the cloned genome during experimental transfections, both *in vitro* and *in vivo* (Jourdan *et al.*, 1990).

Hairpin structure of ambidensoviruses, especially for those classical viruses belonging to the *Lepidopteran ambidensovirus* 1 group possess relatively long ITRs. JcDNV has ITRs of 517 nts long (Dumas *et al.*, 1992), which exceeds the size of all

known parvovirus ITRs, such as B19 (383 nts) (Zhi *et al.*, 2004) and AAV2 (145 nts) (Lusby *et al.*, 1980). The distal 96 nucleotide form a Y-shaped hairpin which is typical for parvovirus termini with the flip and its complement flop configuration. Their terminal structure contain at least four important elements: (i) the origin of replication at the approximate position of nt 95; (ii) a site-specific nicking enzyme that recognizes only ds DNA at approximately the same site in the complementary strand (terminal resolution); (iii) a flip/flop region (nts 35 - 65) that is most likely important in positioning enzymes that act at the genome ends, and, (iv) for some members (i.e. GmDNV, JcDNV, MIDNV, PiDNV), promoter elements for both the NS and VP genes (Bergoin *et al.*, 2000). The other non-classical ambidensoviruses have shorter ITRs with a different structure. For AdDNV, the ITRs are 144 nts, of which the distal 114 nts could fold into a perfect I-type palindromic hairpin (Liu *et al.*, 2011).

The iteradenovirus genomes have ITRs that are only 225 - 230 nts long of which the terminal 150 - 160 nts may form a hairpin. This hairpin does not form a typical T- or Y-type structure as seen with the other parvoviruses but possesses a J-shaped structure with two imperfectly base-paired side arms and can be found in either form, flip or flop (Fédière *et al.*, 2002, Li *et al.*, 2001) (Figure 1.6 B). The origin of replication is therefore located about 70 nts into the ITR and the nicking site is predicted to be located at about 165 nts into the ITRs.

The brevidensoviruses have the smallest genomes among all parvoviruses. Their terminal structures lack ITRs and resemble the majority of vertebrate autonomous parvoviruses. This dissimilarity at the ends may also explain why different amounts of the complementary strands are encapsidated. This may be due to a difference in the packaging signal (probably in the ITRs, which in the other two genera result in equimolar packaging) or due to asymmetric strand displacement (Bergoin *et al.*, 2000). However, these terminal hairpins can fold back on themselves into T-shaped structure with 134 nts and 182 nts at either ends (Figure 1.6 C).

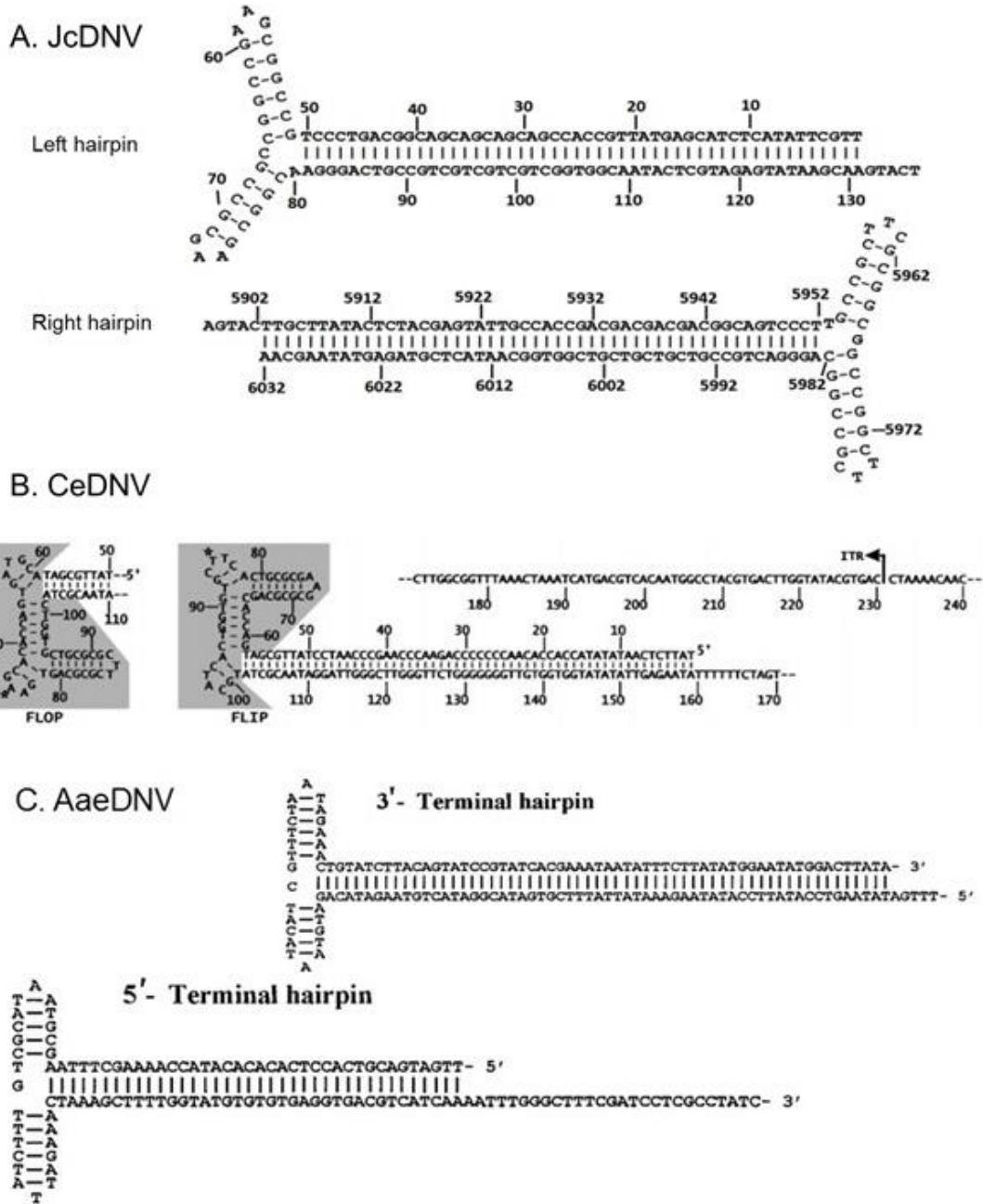


Figure 1.6 Hairpin structure of different densoviruses.

A) JcDNV (*Ambidensovirus* genus) left and right hairpins(Pham et al., 2013d). **B)** CeDNV (*Iteradensovirus* genus) flip and flop (Fédière et al., 2002). **C)** AaeDNV (*Brevidensovirus* genus) 3' and 5' terminal hairpins. (Bergoin et al., 2000, Fédière et al., 2002)

2.6 Viral Proteins

Parvoviruses are among the smallest known viruses. They contain one copy of a nonpermuted, linear, single-stranded DNA genome, ~5 kb in length. The coding sequence is flanked by short (116–385-base) imperfect palindromes that can fold into hairpin structures. The terminal hairpins bracket a single-stranded coding region of 4.5–5 kb that contains two major gene cassettes, the 5' half of the genome encodes a small number of nonstructural (NS) proteins essential for replication, and the 3' half encodes two or more N-terminal variants of a single capsid protein (Cotmore *et al.*, 2014b).

2.6.1 Nonstructural proteins and rolling cycle replication

Parvovirus genome replication relies on the host cell's DNA polymerase during S-phase, using the terminal hairpins of the single-stranded genome as *cis* primers, to generate double-stranded DNA and a rolling circle-like mechanism initiated by a distinct virus-encoded endonuclease for genome amplification (Cotmore *et al.*, 2006b). The rolling-circle replication initiation endonuclease (RCRE) and the superfamily 3 helicase (SF3H) that are involved in this process are located on the multifunctional, regulatory non-structural 1 (NS1) protein (Nüesch, 2006). NS1 is the most conserved protein among vertebrate and invertebrate parvoviruses since it contains two important motifs; rolling circle replication (RCR) motif and a Superfamily 3 (SF3) helicase motif (Dumas *et al.*, 1992, Fédière *et al.*, 2004, Tijssen *et al.*, 2003) (Figure 1.7). The gene for NS2, another Rep protein implicated in the subcellular trafficking of viral capsid proteins (Cotmore *et al.*, 1997), usually overlaps the NS1 gene, either by (i) sharing the N-terminus and connecting to a different ORF by splicing (Bergeron *et al.*, 1993, Schoborg *et al.*, 1991); by (ii) having an ORF in a different frame (overlapping promoters) (Pham *et al.*, 2013a); or (iii) using alternative splicing or leaky scanning (Tijssen *et al.*, 2003). The RCRE domain is located within the N-terminus of NS1, whereas the SF3H domain is usually found near the middle of the protein (Astell *et al.*, 1987, Nüesch, 2006). The C-terminal domain of NS1 has different activities, such as

the transactivation of the promoter for the capsid protein (Legendre *et al.*, 1992, Legendre *et al.*, 1994, Rhode *et al.*, 1987). The NS1 protein of the ambisense densoviruses is about 200 amino acids shorter than that of other parvoviruses or densoviruses (Tijssen *et al.*, 2006a). However, this is compensated by an extra NS3 protein coded upstream of the NS1 and NS2 genes. These NS3 proteins do not form a monophyletic clade but are closely related to a protein of granulosis viruses (Baculoviridae) (Hörer *et al.*, 1995, Lange *et al.*, 2003, Tijssen *et al.*, 2006a). These proteins share with AAV Rep78/52 proteins the presence of Zn-finger motifs (C-X(2)-C). It is tempting to speculate that NS3 could be an apoptosis inhibitor and play a role in overcoming the antiviral state in the infected cell (Crook *et al.*, 1993).

	RCR motif		ATPase motif	
	motif-2	motif 3	Walker A-site	B-site
JcDNV	132 GDHIHVIHD-41-DVFIYFFVRKR		403 FLIISPPSAGKNFFFDMIFGL-25-VLLWNEPNYE	
GmDNV	131 GDHIHVIHD-41-DVFIYFFVRKR		402 FLVMSPPSAGKNFFFDMIFGL-25-VLLWNEPNYE	
BmDNV-1	289 QGHFHILHA-36-NIMFYNTKWPR		560 FQIVSPPSAGKNFFIETVLAF-25-VNYWDEPNFE	
CeDNV	289 EGHFHILHA-36-NIMFYNTKWPR		559 FQIVSPPSAGKNFFIETVLAF-25-VNYWDEPNFE	
Aa1DNV-1	316 GDHIHILFS-35-NYILYCIRYGI		579 MVLEGITNAGKSLILDNLAM-24-SVLFEPMIT	
AaeDNV	372 GDHIHILFS-35-NYILYCIRYGI		620 MVLEGITNAGKSLILDNLAM-24-SILFEPMIT	
MVM	125 GWHCHVLIG-72-MIAYYFLTKKK		395 VLFHGPASTGKSIIAQAAIAQA-23-LIWVEEAGNF	
PPV	126 GYCHCHVLLG-69-MIAYYFLNKRR		394 ILFHPGASTGKSIIAQHIANL-23-LIWIEEAGNF	
AAV2	88 YFHMHVLLM-55-YIPNYLLPKTQ		330 IWLFGPATTGKTNUAEAIHT-23-VIWWEEGKMT	
SAAV	77 GYHMHVLLN-54-YLKNYFRKTL		324 IWLYGPATTGKTIIAQAAIAHA-23-LIWWEEGKMT	
B19	79 GYHIHVVIG-50-FIENYLMKKIP		324 LWFYGPSTGKTNLAMAIKS-23-LVVWDEGIIK	
SiPV	79 GFHIHVVIG-49-FVTYYLMPKLY		320 IWLFGPSTGKTNIAMSLASA-23-IILWDEGLIK	
consensus	.uHuHuuu.	.u..Yu..K..	uuu.GP...GKTu.....	uu..DD....EE
		R	S	S

Figure 1.7 Conserved regions between densoviruses and vertebrate parvoviruses.
Conserved rolling circle replication (RCR) and Superfamily 3 (SF3) helicase motifs found in vertebrate and invertebrate parvovirus. Conserved residues in Ambidensovirus and Iteradensovirus, the Walker A-site, contain GKN residues instead of GKS/T. Ambidensovirus genus members usually have a single E residue in the conserved Walker B-site rather than double residues in other parvoviruses (Tijssen *et al.*, 2006a).

2.6.1.1 Rolling hairpin replication

The single-stranded DNA genome of the virus needs to be converted to a double-stranded intermediate in order to start transcription and translation of the viral genome and proteins, respectively (Berns, 1990). Densoviruses have received very little

attention regarding their mode of replication but the structural similarities of their genomes with that of vertebrate parvoviruses suggest common strategies in the replication of the DNA, strand displacement and encapsidation (Bergoin *et al.*, 2000). Replication is active during the host cell S phase and occurs in distinct stages. The DNA polymerase δ , the most probable cellular enzyme responsible for the synthesis of the complementary strand, is expressed in cells only during the S phase of the cell cycle (Cossons *et al.*, 1996). Rolling-hairpin replication (RHR) is a linear adaptation of the rolling-circle replication (RCR) mechanisms used by many small circular replicons and is mediated predominantly by a subset of the synthetic machinery of the host, supplemented and orchestrated by the viral initiator protein, NS1 (Cotmore *et al.*, 2014b). The viral endonuclease NS1 creates a nick, the nickase is left covalently attached to the 5' end of the viral DNA via its active-site tyrosine, where it persists throughout replication, packaging, and virion release (Cotmore *et al.*, 2014b). Within host cells, all RCR replicons amplify through a series of duplex replicative-form (RF) DNA intermediates, and in general employ distinctive origins of replication that contain small palindromic DNA sequences, which are able to alternate between inter- and intrastrand basepairing at different phases of the replication process (Cotmore *et al.*, 2006b). During the amplification phase, a single continuous DNA strand, synthesized by a unidirectional fork driven by DNA polymerase δ , creates a series of concatemeric duplex intermediates (Berns *et al.*, 2013, Cotmore *et al.*, 2006c), as illustrated in Figure 1.8.

In step i the left telomere of incoming virion DNA folds back on itself, allowing the base-paired 3' nucleotide to prime synthesis of a complementary positive-sense strand. This creates a monomer-length duplex intermediate in which the two strands are joined at their left end via a single copy of the 3' hairpin, creating a turnaround form of the terminus. This duplex functions as a transcription template for NS1 expression, which is essential for all further steps in the replication pathway because the cellular machinery appears unable to melt and copy the viral 5' hairpin. The two hairpin telomeres play pivotal roles in the ensuing rolling mechanism, acting as hinges that first unfold, allowing themselves to be copied to generate an extended-form duplex

copy of the entire hairpin (steps ii and iii). Because parvoviral hairpins are imperfect palindromes, the new flop sequence (*r*) can be recognized as the inverted complement of its flip template (*R*). This extended-form duplex is then melted and refolded into two hairpins (step iv), creating a rabbit-ear structure that pairs the 3' nucleotide of the newly synthesized DNA with an internal base, effectively reversing the path of the fork and redirecting it back along the internal coding sequences (step v). Both unfolding and refolding the hairpins require NS1, which binds site specifically to duplex motifs in the telomere and requires a functional SF3 helicase domain to assist in melting the duplex (Lou *et al.*, 2004, Willwand *et al.*, 2002). The fork then progresses back along the monomeric duplex, displacing the original negative strand and replacing it with a covalently continuous new strand. During fork progression, the turnaround form of the left hairpin is unfolded and copied, creating the dimer-junction sequence and leading first to the synthesis of a duplex dimer (step vi), which is then similarly processed (step vii) to a tetramer intermediate (step viii). In newly released virions, the 5' end of the genome remains covalently attached to a copy of the viral replication initiator protein (called NS1 in most genera, but Rep1 or Rep68/78 in the dependoviruses), which is located on the outside of the particle, attached to the protected bulk of the genome via a short 'tether' nucleotide sequence that projects through the capsid wall (Cotmore *et al.*, 1989). This sequence, with its attached NS1, can be removed without impairing particle infectivity, and is generally cleaved during natural infection as the virus transits the extracellular environment or is trafficked through the endosomes of its prospective host cell (Cotmore *et al.*, 2006a). Overall, the result of rolling hairpin synthesis is that the coding sequences of the virus are copied twice as often as the termini, and duplex dimeric and tetrameric concatemers are generated, in which alternating unit-length viral genomes are fused, through a single palindromic junction, in either a left-end:left-end or right-end:right-end configuration.

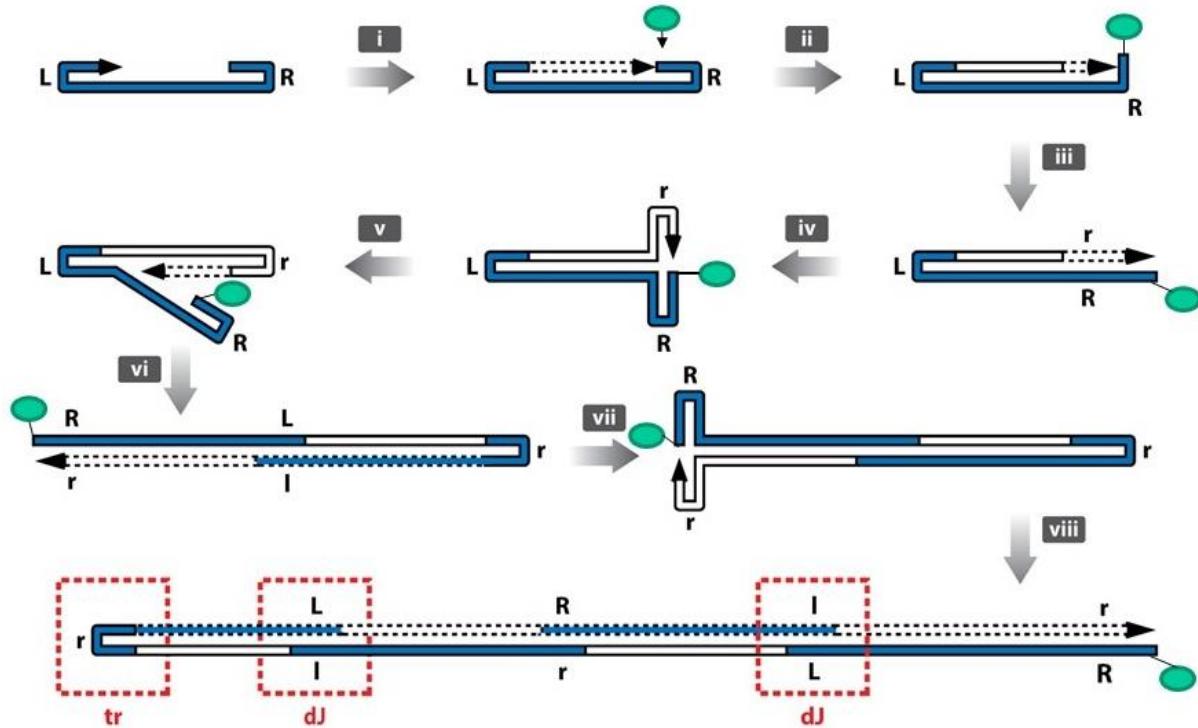


Figure 1.8 Parvoviral Rolling-hairpin replication (RHR) scheme.

The parvoviral genome is represented by a continuous line, blue lines for the original genome, and white lines for progeny genomes. L and I represent flip and flop forms of the left-end telomere; R and r represent these sequence arrangements of the right-end telomere. Newly synthesized DNA is represented by dashed lines, and its 3' end is capped with an arrowhead. The green sphere represents an NS1 molecule. Red dotted boxes designate the turnaround (tr) form of the right end and the dimer junction (dJ) form of the left-end palindrome (Cotmore *et al.*, 2014b).

2.6.2 Structural proteins

Theoretically, structural proteins of 25-30 kDa would suffice to form a capsid to protect the viral genome, according to the atomic structure of parvoviruses. However, this minimum size is often greatly exceeded. All the parvoviruses contain two to five different proteins, which usually have molecular masses between 40 and 100 KDa, depending on the species. The different structural proteins for each virus share common sequences and can be distinguished by electrophoresis through polyacrylamide gels in the presence of sodium dodecyl sulphate (SDS-PAGE) (Tijssen *et al.*, 1981). The production of these N-terminal isoforms of viral proteins is a strategy employed by densovirus to maximise the coding capacity of their genomes which are limited to 4-6 kbs. About half of the genome is dedicated to produce the 4 structural proteins coded by one gene.

The classical members in *Ambidensovirus* genus such as GmDNV (Tijssen *et al.*, 1976), JcDNV (Dumas *et al.*, 1992), MIDNV (Fédière *et al.*, 1995) and PiDNV (Huynh *et al.*, 2012) share about 80-90% sequence identities and they have strong serological cross-reactivities among their structural proteins. They have four forms of structural proteins produced from a unique open reading frame (ORF), with molecular masses of about 47, 53, 58 and 89 KDa. The four proteins are usually found in different ratios within the viral particle. The ratio between VP1, VP2, VP3 and VP4 in GmDNV is 2:15:15:68 (Tijssen *et al.*, 1977, Tijssen *et al.*, 1976) and in JcDNV is 14:26:21:39 (Fédière, 1983). VP1 is a minor protein translated from the whole mRNA transcript. The smallest VP4 is the most abundant and can form a virion or the virus-like particle by itself when expressed in a baculovirus system (Croizier *et al.*, 2000).

Similar to vertebrate parvovirus NS1, the densovirus NS1 proteins also contain a replication initiator, ATP binding and the helicase superfamily III motifs. However, the position of these elements in the densovirus NS1 molecules can be quite different (Bergoin *et al.*, 2000) (Figure 1.9 A). Densoviruses share with vertebrate parvoviruses

the property that their VP gene generates a nested set of proteins with identical C-terminal sequences. However, there are only two conserved motifs within the VP coding sequence. The first conserved motif is found in the N-terminal region of the VP2 initiation codon of vertebrate parvoviruses and VP4 initiation codon of densoviruses. This motif is an S/T-G rich stretch and is believed to be involved in the N-terminal delivery of VP1 to the surface of the viral capsid through the five-fold channel (Simpson *et al.*, 1998, Tsao *et al.*, 1991) during cell entry. The second conserved motif is found in the N-terminus of VP1 unique part (VP1up) of both densoviruses and vertebrate parvoviruses. This region contains a highly conserved PGY motif (Figure 1.9 A), which was later shown to have a phospholipase A2 (PLA2) enzymatic activity (Zádori *et al.*, 2001). Some amino acids in the conserved domain of VP1up match critical amino acids in the catalytic site of secretory phospholipase A2 (sPLA2), an activity that was not known to exist in virus capsids. This PLA2 domain resides within the N-terminal region of VP1up of almost all known parvoviruses and is required for successful infection (Zádori *et al.*, 2001). The Ca²⁺ binding loop (GPGN) as well as the catalytic site of PLA2 (DxxAxxHDxxY) were found to be conserved among both vertebrate and invertebrate parvoviruses (Figure 1.9) (Tijssen *et al.*, 2006b). The conservation between VP1up and sPLA2 is poor outside the HDxxY and YxGxG motifs (Figure 1.9 B). Brevidensovirus, hepadensovirus and Aleutian mink disease virus (AMDV) are the only parvoviruses that do not have PLA2 motifs. The PLA2 activity of different parvoviruses could differ by as much as 1000-fold, with PPV being as active as the most active sPLA2 (Tijssen *et al.*, 2006b). Viral infectivity of PPV decreases dramatically when amino acids in the HD catalytic dyad are mutated. Moreover, the specific activity of the PLA2 of *Lepidopteran Ambidensovirus 1* members was significantly lower than that of the *Iteradensovirus* or vertebrate parvoviruses (Fédière *et al.*, 2004, Fédière *et al.*, 2002, Li *et al.*, 2001, Tijssen *et al.*, 2003).

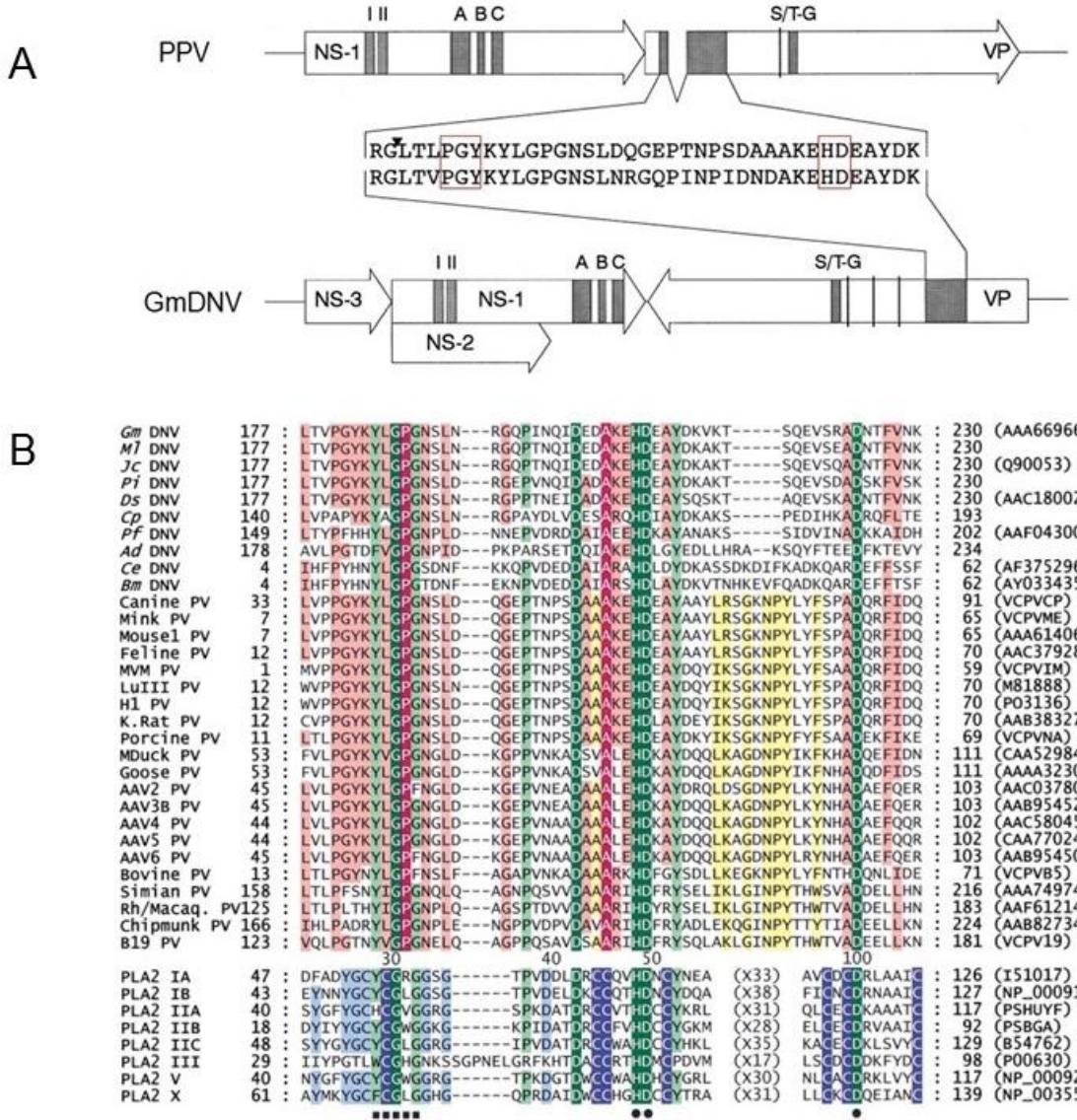


Figure 1.9 Conserved motifs within the viral genome and PLA2 motif.

A) Localization of the conserved motif on the viral genome of vertebrate parvovirus PPV and Densovirus GmDNV. B) Sequence Alignments of Parvovirus PLA₂ Motifs and sPLA₂ representatives. Insect parvoviruses are indicated by DNV and their host name: Gm, *Galleria mellonella*; MI, *Mythimna loreyi*; Jc, *Junonia coenia*; Pi, *Pseudoplusia includens*; Ds, *Diatraea saccharalis*; Cp, *Culex pipiens*; Ce, *Casphalia extranea*; Ad, *Acheta domesticus*; Pf, *Periplaneta fuliginosa*; Bm, *Bombyx mori*. Vertebrate parvoviruses are indicated with PV. K. Rat, kilham rat; MVM, minute virus of mice; Mduck, Muscovy duck; Rh/Macaq., rhesus-macaque monkeys; AAV, adeno-assotiated virus. Lulll and H1 parvoviruses originated from tissue cultures, and B19 is a human parvovirus. IA: *Naja naja* snake venom, IB: human pancreatic, IIA: human synovial fluid, IIB: gaboon viper snake venom, IIC: rat, III: bee venom, V: human PLA2, X: human. The Ca²⁺ binding loop is underscored by filled squares and the catalytic residues by filled circles. Dark green background (white lettering), 100% identity; light green background, at least 70% conserved among all PLA2s. Dark and light red indicate 100% identity and at least 70% conserved, respectively. Yellow background indicates conserved residues in vertebrate parvoviruses [conserved amino acids: (L, V, I), (Y, W, F), (R, K)] (Zádori et al., 2001).

3 Atomic structures of viral particles

Parvoviruses package a single copy of their non-permuted, single-stranded, linear genome into robust $T = 1$ icosahedral protein capsids 18–26 nm in diameter, which lack histones or accessory proteins (Cotmore *et al.*, 2013). The protein capsid provides a protective coat to the DNA as it encounters environmental challenges in transmission from host to host, and cell to cell. The capsid may have several other functions: recognition of appropriate host cells, entry, intracellular transport, release of nucleic acid at the appropriate time and place, assembly of progeny virus and virion release (Chapman *et al.*, 2006). At atomic resolution, parvoviral capsid structures have been the exclusive domain of X-ray crystallography. Since 1991, the crystal structures of particles from several vertebrate parvoviruses and densoviruses have been determined, including canine parvovirus (CPV) (Tsao *et al.*, 1991), B19 (Agbandje *et al.*, 1994, Kaufmann *et al.*, 2004), minute virus of mice (MVM) (Agbandje *et al.*, 1998), FPV (Agbandje *et al.*, 1993), PPV (Simpson *et al.*, 2002), AAV2 (Xie *et al.*, 2002), GmDNV (Simpson *et al.*, 1998), BmDNV (Kaufmann *et al.*, 2011) and PstDNV (Kaufmann *et al.*, 2010).

In each case, the modeled $T = 1$ icosahedral capsid contains 60 highly interdigitated copies of the VP core domain. This sequence establishes two-, three- and fivefold symmetry-related interactions (Figure 1.10 A), which define the icosahedral asymmetric unit of the particle and create its characteristic topological features (Cotmore *et al.*, 2014b). The surface can be divided into an icosadeltahedron of equal triangles (monomers), each designating a common protein, usually the major protein (Figure 1.10 A). The major protein is VP2 in CPV, VP4 for GmDNV, and VP3 for BmDNV-1. For members of the genus *Protoparvovirus* the proportion of minority proteins (VP1) and the majority (VP2) is between 1: 5 and 1:10. This ratio is controlled by producing respective mRNA. The alternative splicing mechanism makes VP2 mRNA production more efficient than VP1 mRNA production (Tullis *et al.*, 1993).

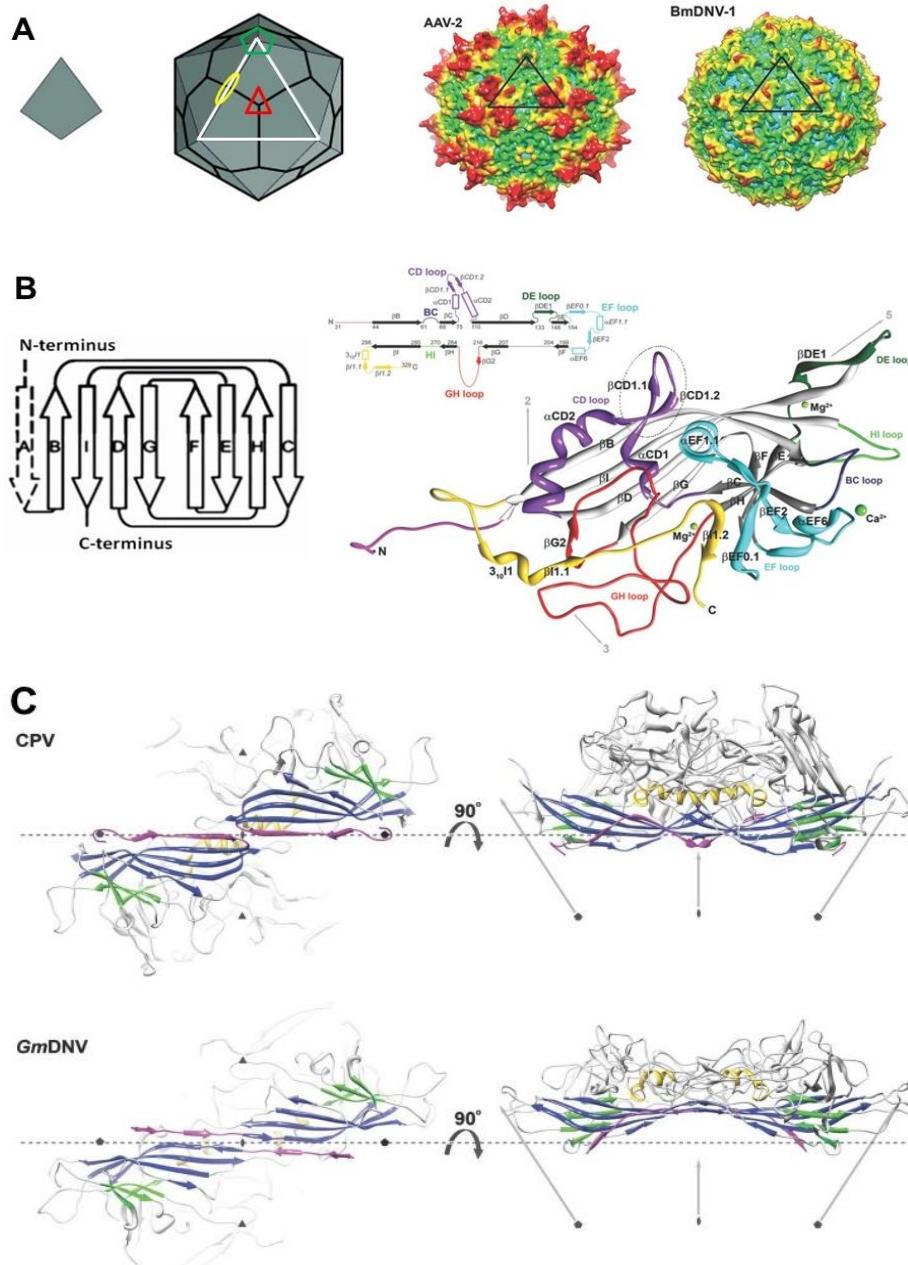


Figure 1.10 The architecture of parvovirus.

A) The structure of $T = 1$ is composed of 60 copies of the structural proteins (adapted from viralzone.expasy.org). In parvovirus these proteins assemble into first trimer (white triangle). The 3-fold axis is at the center of the trimer (red triangle). The side of the trimer represent the 2-fold axis (yellow oval), and the tip of five trimers form the 5-fold axis (green pentagon). The capsid surface topology of AAV-2 and BmDNV-1 represent the trimers. B) The topology of each subunit is a viral jellyroll β -barrel (left) and the 3D-fold of PstDNV subunit (Kaufmann et al., 2010). C) Spatial arrangement of the parvoviral core jelly roll and the N-terminal region for the capsid proteins of densovirus GmDNV and vertebrate parvovirus CPV. Conserved secondary structure elements of each protein subunit are colored blue (β -BIDG), green (β -CHEF), and gold (helical elements). The N-terminal region of the capsid protein, upstream of β B, including β A, is shown in magenta. The positions of icosahedral symmetry axes are indicated by polygonal symbols and arrows.

Capsid protein sequences of parvoviruses share very low sequence identities but the topology at atomic resolution is often quite similar. Each subunit contains a “jelly roll” or β -barrel motif in which there are eight anti-parallel β -strands alphabetically named from N- to C-terminus (Figure 1.10 B). This β -barrel structure is highly conserved in many viral families (Nandhagopal *et al.*, 2002, Rossmann *et al.*, 1989). By convention, loops are named according to the strands between which they bridge, so that the BC loop is between strands B and C. The loops of different viruses have little in common, other than in general terms: loops BC, HI, DE and FG, which are close to the 5-fold axis, tend to be shorter than the CD, EF, and GH loops (Chapman *et al.*, 2006). Eight β -strands are arranged into two four-stranded sheets, BIDG and CHEF. The BIDG sheet is located near the inner surface of the capsid and is more regular and contains longer strands than the CHEF sheet. In all parvoviruses, the BIDG sheet is 5-stranded with an N-terminal edge strand β A. There are some differences between vertebrate parvoviruses and invertebrate densoviruses. For example, all of the parvoviruses contain a tight U-turn between β B and β A, but in densoviruses the β A is an N-terminal extension of β B to become the first strand of the BIDG sheet of a 2-fold related subunit (Simpson *et al.*, 1998). This exchange between two subunits creates a “swapping domain”, a common feature found in insect viruses (Kaufmann *et al.*, 2010), suggesting the domain swapping is an evolutionarily conserved structural feature of the *Densovirinae* (Kaufmann *et al.*, 2011). In invertebrate parvoviruses, the GmDNV capsid was the first to be resolved at the near-atomic structure (3.7 Å resolution) making it possible to compare vertebrate and invertebrate parvoviruses (Simpson *et al.*, 1998). The comparison between the structures of densovirus GmDNV and canine parvovirus (CPV) shows that the central structure β -sheets are highly similar between the two viruses, while loops are highly different. In the vertebrate parvovirus CPV, the β -BIDG sheet is extended to an anti-parallel β -ABIDG sheet by the backfolded β -strand A (β A) that is located in the N-terminal portion of the protein sequentially upstream of β B. In contrast, in the densovirus GmDNV, β B is essentially the linear extension of β A (Figure 1.10 C). β -barrels are a common feature of viral proteins in both densoviruses and vertebrate parvoviruses. However, the β -barrel of GmDNV is rotated by 7.4 Å and translated

radially inwards by 9.7 Å to superimpose it on the β-barrel of the capsid protein of CPV, when the rotational symmetry axes are superimposed (Simpson *et al.*, 1998).

Parvoviral VP3 is about three times the size of the smallest jellyrolls from other viruses, this is due to the long loops between the strands (Chapman *et al.*, 2006). In most parvovirus structures, only about 20-30% of the protein mass is contained in the β-strands while the remaining, about 70-80%, make up the loops connecting the strands. Unlike the β-strands, loops are known to differ among parvoviruses and account for surface structure, especially at the 3-fold axes. Variable capsid surface loops govern many biological functions, including receptor binding, tissue tropism, pathogenicity, and antigenicity; and, also form the structural basis of the phenotypic differences among members and strains of the same virus (Agbandje *et al.*, 2006). In BmDNV-1, the GH loop is more intertwined with its 3-fold-related neighboring subunit than in GmDNV, but both are in contrast to PstDNV, for which this loop does not interdigitate between neighboring subunits (Kaufmann *et al.*, 2010). The structure and length of the loop regions in BmDNV-1 show some similarity to GmDNV, but they differ greatly from the vertebrate parvoviruses, as well as the shrimp parvovirus PstDNV (Figure 1.11), suggesting a smaller divergence in evolutionary development between BmDNV-1 and GmDNV than between these and the vertebrate parvoviruses, as well as PstDNV (Kaufmann *et al.*, 2011).

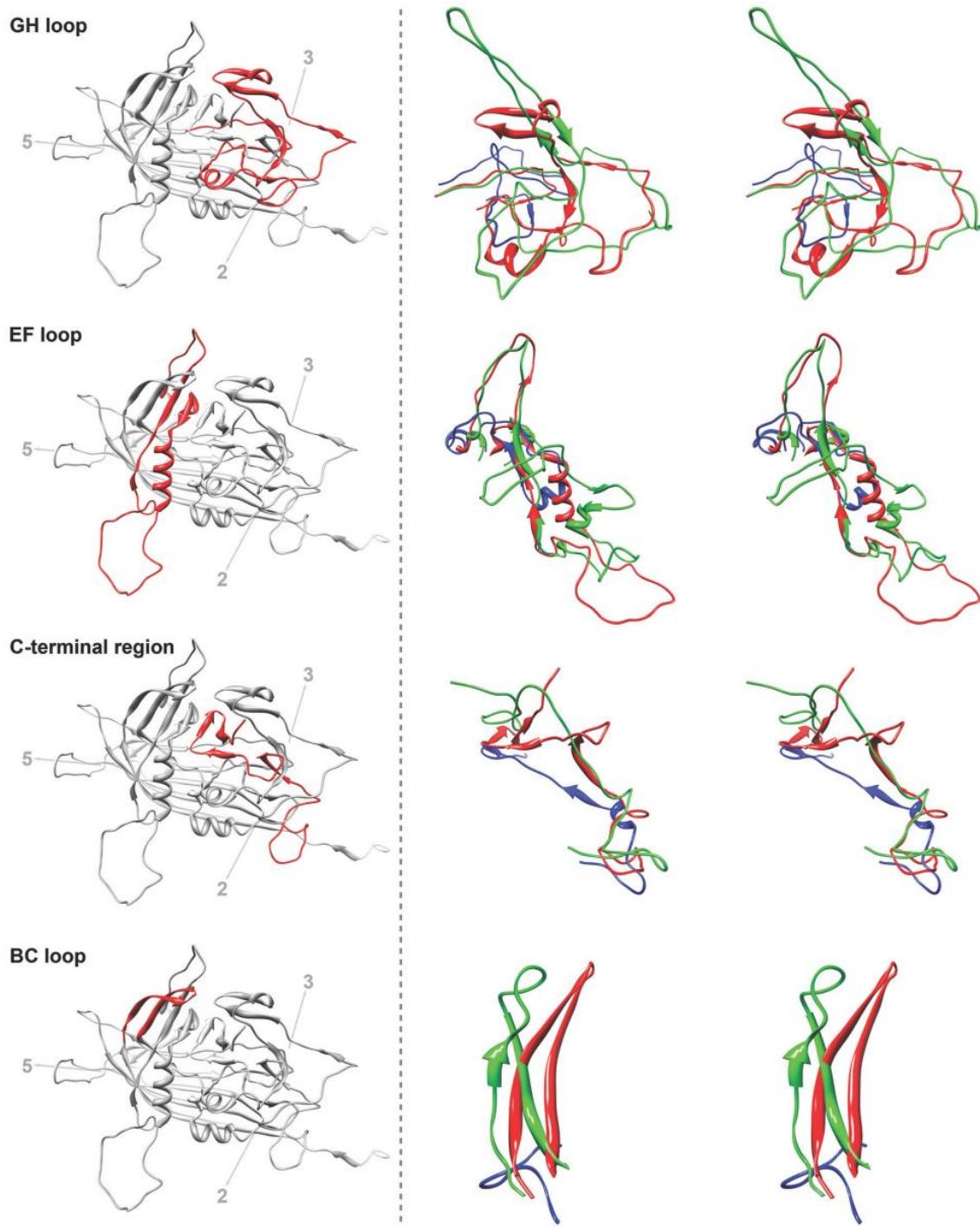


Figure 1.11 Structure comparison of three densoviruses.

Structure comparison of loops GH, EF, BC, and the C-terminal region of BmDNV-1 with those of GmDNV and PstDNV. Left, a ribbon diagram of the BmDNV-1 capsid protein indicates the position of the particular loop (red) relative to the icosahedral symmetry axes. Right, close-up stereo views of the superpositioned loop regions of BmDNV-1 (red), GmDNV (green), and PstDNV (blue) (Kaufmann et al., 2011).

Adeno-associated viruses (AAVs) are small, nonenveloped, single-stranded DNA (ssDNA) viruses that belong to the *Dependovirus* genus of the *Parvoviridae* family (Muzyczka *et al.*, 2001). Currently, 13 distinct human and nonhuman primate AAV serotypes (AAV1–AAV13) have been sequenced and PCR studies of both nonhuman primate and human tissues have identified numerous other AAV genomes (Drouin *et al.*, 2013). In the past few decades, recombinant AAVs have become promising vectors for therapeutic gene delivery due to their ability to package and express foreign genes in the absence of active cell division in a broad range of tissues and without any associated pathogenicity (Berns *et al.*, 1995, Erles *et al.*, 1999). The AAV serotypes demonstrate unique tissue tropisms, as well as other characteristics including differential transduction efficiencies and blood-clearance properties (Asokan *et al.*, 2010, Asokan *et al.*, 2012, Kotchey *et al.*, 2011, Mingozi *et al.*, 2011). These unique properties are likely dictated by their viral capsid structure. To date, the capsid structures of AAV1–AAV9 have been determined.

In an assembled capsid, the VP monomers interact at icosahedral (two-, three- and five-fold) symmetry axes (Figure 1.12 B) which is similar to all the parvoviruses. The 2-fold axis region is believed to be responsible for conformational changes in the capsid that arise during viral endosomal trafficking (Nam *et al.*, 2011). The 3-fold axis region has been shown to be important for receptor binding and antibody recognition (Gurda *et al.*, 2013, Opie *et al.*, 2003). The AAV VP3 structure contains highly conserved regions that are common to all serotypes, a core eight-stranded β -barrel motif (β B- β I) and a small α -helix (α A) (Drouin *et al.*, 2013) (Figure 1.12 A). The loop regions inserted between the β -strands consist of the distinctive HI loop between β -strands H and I, the DE loop between β -strands D and E, and nine variable regions (VRs), which form the top of the loops. These VRs are found on the capsid surface and can be associated with specific functional roles in the AAV life cycle including receptor binding, transduction and antigenic specificity. In general, learning the structural differences between the serotypes can be exploited for designing tropisms to specific tissues or cell types for the treatment of specific diseases.

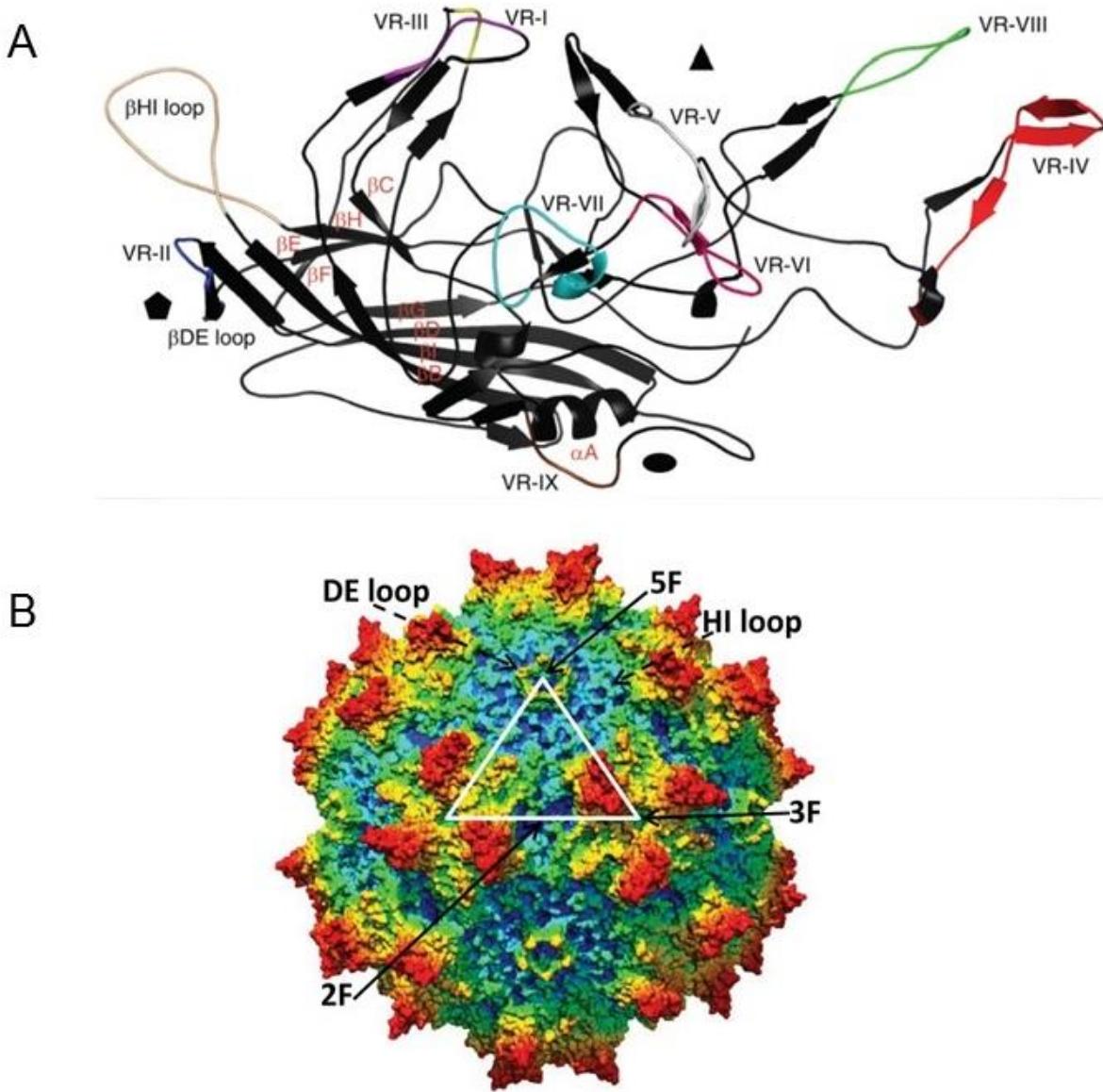


Figure 1.12 Adeno-associated virus (AAV) capsid structure.

A AAV VP3 monomer is shown with a conserved core region consisting of eight antiparallel β -sheets (β B– β I) and an α -helix (α A). Loop insertions between the β -sheets vary among the AAV serotypes. Nine variable regions (VRs) are coded and noted (I: purple; II: blue; III: yellow; IV: red; V: gray; VI: hot pink; VII: cyan; VIII: green; IX: brown; β HI loop: tan) (Drouin *et al.*, 2013). B Surface topology structure of AAV1. Radially color-cued (from capsid center to surface, blue to green to yellow to red) surface representation of the AAV1 capsid. The white triangle depicts a viral asymmetric unit bounded by one 5-fold axis and two 3-fold axes with a 2-fold axis between them. The approximate locations of the icosahedral 2-fold (2F), 3F, and 5F axes are indicated by the black arrows. The positions of the DE and HI loops are indicated by the dashed arrows (Nam *et al.*, 2011).

Chapter II: Materials and Methods

Cells and medium

Four established insect cell lines were used in this study: (I) Sf9 cells derived from *Spodoptera frugiperda* ovaries (Vaughn et al., 1977) cultured in Sf-900 II medium supplemented with 5% (v/v) heat-inactivated FBS (foetal bovine serum) and 1% (v/v) antibiotic mixture of penicillin (50U/ml), streptomycin (5mg/ml). (II) LD652 cells from *Lymantria dispar* ovaries (Goodwin, 1985), also maintained in Sf-900 II with 5% FBS and antibiotic mixture. (III) BmN cells from *Bombyx mori* also maintained in Sf-900 II with 5% FBS and antibiotic mixture. (IV) C6/36 cells from *Aedes albopictus* cultured in RPMI medium supplemented with 10% (v/v) FBS and 1% (v/v) antibiotic mixture.

All the insect cell lines were incubated at $28 \pm 1^\circ\text{C}$ and regularly passaged by pipetting before they reached 100% confluence (usually around 4 to 5 days).

Insect

In order to identify the 5' and 3'-ends of viral transcripts, infected *Papilio* larvae were used for isolating viral mRNA. The *Papilio* eggs were bought from a farm in Granby, Quebec. *Papilio* larvae were fed with pesticide-free parsley leaves during the summer time (room temperature around 25°C).

Cloning vectors and competent cells

Five vectors were used for different cloning purposes: (I) pGEM-T & pGEM-Teasy (Promega) vectors were used for TA-cloning of PCR products, such as the SISPA sequence analysis and 3'/5'-RACE products sequences analysis. (II) pBluescript SK(±) were used for infectious clone construction of SfDV and DppIDV. (III) The PCR2.1-60 vector which was derived from the TA cloning vector PCR2.1-TOPO by inserting a *Bam*H site were used for constructing the infectious clone of PpDV. (IV)

pIZT/V5-His (Invitrogen), carrying the green fluorescent protein (GFP) and OptE promoters was used to express genes in insect cell lines (Sf9 and LD652).

During the current work, two different competent cells were used for cloning construct transformation: *E. coli* 10G cells (Lucigen) for most of the TA-cloning and pIZT/V5-cloning; *Sure* cells for all the infectious clones (including hairpins) construction. All cells were prepared for electroporation. *E. coli* 10G was incubated at 37°C and *Sure* cells were always incubated at 30°C.

Sequence-independent single primer amplification (SISPA).

The SISPA protocol according to Allander (Allander *et al.*, 2001) was adapted as follows for iteradenovirus isolation. Approximately 0.5g of *Papilio polyxenes* and *Sibine fusca* larvae or *Danaus plexippus plexippus* pupae were each mixed with 800 µl PBS and incubated overnight at 4°C. Then 200 µl carbon tetrachloride was added, mixed well with the sample and centrifuged at high speed to separate phases. One hundred units of DNaseI was added to 200 µl of each upper phase to digest the contaminating (host) DNA. Subsequently, viral DNA was extracted with a High Pure Viral DNA Isolation Kit (HPVDI kit, Roche). Second-strand DNA was synthesized from all extracted DNA with 10 pmol of random hexamers and 5 units of 3'-5' exo-Klenow fragment of DNA polymerase (New England Biolabs) for 1 h at 37°C in 60 µl of a buffer containing 200 nM each dNTP, 10 mM Tris-HCl, pH 7.5, 5 mM MgCl₂, and 7.5 mM DTT. The second-strand synthesis reaction was terminated by heat-inactivation at 72 °C for 10 min and the duplex DNA was digested with 10 units of Csp6.I (Fermentas, Hanover, MD) for 1 h at 37°C (Allander *et al.*, 2001). Then, the DNA fragments were extracted, precipitated, dissolved in H₂O, and ligated to the adaptor NCsp [hybridized oligonucleotides NBam24, AGG CAA CTG TGC TAT CCG AGG GAG; and NCsp11, TAC TCC CTC GG) in a 10 µl reaction containing 20 pmol of adaptor and 6 units of T4 DNA ligase (New England Biolabs) in a standard ligation buffer at 4°C overnight (Allander *et al.*, 2001). For PCR, 2.5 µl of the ligation reaction was used as a template

in the 50 μ l reaction mix and 250 ng of NBam24 were used as primer. Before adding 2.5 units of Taq polymerase, the PCR reaction was heated to 72°C for 3 min. Amplification was completed in a thermal cycler under the following conditions: 1 cycle at 94°C for 3 min, 35 cycles of 94°C for 30 s, 55°C for 30s, 72°C for 30 s, followed by an elongation step of 72°C for 5 min.

SISPA amplicon sequencing and preliminary viral genome characterization.

PCR products were separated on a 1.5% agarose gel and bands were cut out directly. DNA was extracted from the gel and cloned into a pGEM-T easy vector. The sequences obtained from different sizes of inserts were analyzed for nucleotide sequence identity (blastn) (<http://www.ncbi.nlm.nih.gov/blast/>) by searching the standard non-redundant databases as well as the high throughput genomic sequences database. Based on the sequences, some gene-specific primers for the three new unknown viruses were designed (Figure 2.1). Together with a highly conserved primer (Pa140) from iteradensovirus ITRs, the unknown genomes were amplified between their ITRs by PCR. For PpDNV, a unique SacI restriction site was present in the middle of the genome. For SfDNV and DppIDV, a single C_{la}I restriction was present in the middle of the genomes so that two halves of the genome could be amplified and joined at these restriction sites.

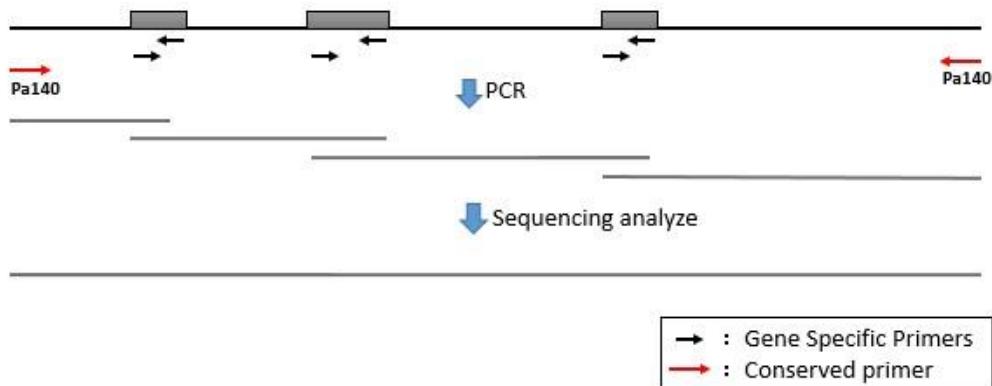


Figure 2.1 Model for identifying new viruses' genome sequence from SISPA results.

Construction of infectious clones.

Virus was purified from larvae or pupae as described previously for GmDNV (Tijssen *et al.*, 2003). Viral DNA was extracted from 400 µl purified virus with the HPVDI kit, according to the supplier's instructions. Treatment of 3.5-4.5 µg viral DNA with 12 units of T4 DNA polymerase and 10 units of Klenow (LG Frag) with 200 µM of the different dNTPs was used for blunt-ending. Then, viral DNAs were digested by the unique-site restriction enzymes. For PpDV, blunt-ended DNA was digested by SacI and cloned into *Eco*RV and SacI sites of PCR2.1-60 vector (derived from the TA cloning vector PCR2.1-TOPO by inserting a *Bam*HI site). Clones with a 1.6-kb insert and clones with a 3.4-kb insert were obtained, and four inserts of each set were sequenced in both directions using Sanger sequencing. Similarly, for SfDV and DppIDV, blunt-ended DNAs were digested by *Cla*I and cloned into *Eco*RV and *Cla*I sites of the pBluescript SK- vector. Clones with a 2.6-kb and a 2.4kb insert were obtained from both SfDV and DppIDV. Combining these inserts into a single vector resulted in infectious clones.

Transfection and infection

Cells transfection: LD652, Sf, Bmn and C6/36 cells were passed in 25 cm² flask or 24-well plate. Before transfection, the cells usually reached 30%-50% confluency. Transfections were done using DOTAP Liposomal Transfection Reagent (Roche) for LD652, Sf, Bmn and C6/36. Four micrograms high pure plasmid DNA (prepared with Qiagen Endofree Midi Kit) and 40 mg DOTAP were used for each 25 cm² flask; 1 µg DNA and 5 mg DOTAP were used for each well of a 24-well plate.

Larvae infection: Parsley leaves were immersed with PpDV virus suspension and fed to third instar larvae.

Isolation of viral RNA.

Total RNA was isolated from *Papilio* larvae 72 h post-infection with PpDV and 48 h post-transfection from LD652 cells transfected with SfDV or DppIDV infectious clones by the NucleoSpin RNA II Kit (Clontech). An extra DNaseI treatment was added after RNA extraction with the Turbo DNA-free kit (Ambion). An RT-negative PCR test was included to verify the absence of DNA. Total RNA extracted was subjected to an mRNA purification using an Poly(A)Purist™ Kit (Ambion).

Northern Blots.

About 20 µg total RNA in a 15 µl volume was added to 30 µl RNA sample buffer (1x MOPS [20 mM morpholinepropanesulfonic acid, 5 mM sodium acetate, 0.5 mM EDTA adjusted to pH 8 with NaOH], 18.5% formaldehyde, 50% formamide), 2 µl loading buffer was added, and the mixture was incubated for 5 min at 65°C and separated by electrophoresis on a 1% (w/v) formaldehyde-agarose gel. Parallel lanes contained RNA size markers (Promega). After migration and washing, RNAs were transferred to positively charged nylon membranes (Roche) by capillary blotting overnight. The blotted membranes were prehybridized with 10 mg/ml herring sperm DNA in 50% formamide before hybridization with 32P-labeled probes. RNA probes for NS and VP gene were transcribed from short parts of NS and VP genes. For this purpose, DNA was amplified by PCR using the primers PpNSRpF700 and PpNSNpR850 for the NS gene, PpVPRpF3200 and PpVPNpR3400 for the VP gene.

Mapping of 5' and 3' ends of viral transcripts.

The most probable locations of the transcripts were predicted from the ORFs obtained by sequence analysis. The transcription maps for RNA isolated from PpDNV-infected *Papilio* larvae or for RNAs isolated from LD-652 cells transfected by infectious

clones for CeDNV, BmDNV-1, SfDNV and DppIDV were established by a FirstChoice RLM RACE kit (Ambion) according to the instructions of the supplier. Primers for RACE are shown in Table 1. All the RACE PCR products were cloned into pGEM-T or pGEM-T easy vectors (Promega) and sequenced by Sanger sequencing. To verify the differences of the 5'-ends of NS1 and NS2 transcripts, specific restriction sites between probable transcription start sites of NS1 and NS2 were used to check the 5'-RACE cloning of NS transcripts.

Expression of non-structural proteins NS1 and NS2.

To study the expression of NS1 and NS2, the pIZT insect cell expression vector was used and a V5-epitope was added to the C-terminus of NS2 and an HA-epitope to the C-terminus of NS1. The fragment from NS1 ATG to NS2 stop codon was amplified by PCR with a CC added to the reverse primer and cloned into the *Kpn*I and *Xba*I cloning sites of the pIZT/V5-His vector, to obtain an in-frame NS2 gene/ V5-epitope construct. The remainder of the NS1 downstream of NS2 stop codon to the NS1 stop codon was amplified by PCR with a HA-tag added to the reverse primer and a mutation in the NS2 stop codon in the forward primer, also changed the NS1 stop codon TAG to a TCT. After PCR, the fragment was cloned into the *Age*I site between V5 and His on the vector of the primary clone (pIZT/V5-SfNS2). Following sub-cloning, the two TAA in the V5 sequence were mutated to CAA and AAA by site-directed mutagenesis with primers pIZ5-V5muF and pIZ5-V5muR. The final construct included the complete NS gene of SfDNV, with the V5 tag in frame with NS2 and HA-tag in frame with NS1. Two separate mutants were created to change the ATG of NS1 and NS2 gene into ACC by the primers shown in Table 1. To test SfDNV NS protein expression driven by the virus' own promoter, the OPIE2 promoter of the vector was replaced by a fragment between the *Bsp*HI and *Spel* sites of the viral genome (from 193nt to 750nt). Two additional mutants were created with the TATA box mutated to GAGA and the CAGT motif mutated to TTGT.

Confocal microscopy.

LD652 cells were cultured in 24-well plate and cells in each well were transfected with 0.5 µg DNA and 5 µl DOTAP reagent. After 48 h post transfection, cell cultures were washed with phosphate-buffered saline (PBS), fixed with 3% (v/v) formaldehyde at room temperature for 30 min and permeabilized with 3% (v/v) Triton at room temperature for 30 min. The primary antibodies were diluted in IF buffer (PBS with 0.01% Tween-20, 0.01% BSA, 0.02% NaN₃), 1:200 of V5 antibody (Invitrogen) for staining of NS2 or 1:50 of HA antibody (Santa Cruz) for staining of NS1. After incubation for 1 h at 37 °C, the cells were washed 3 times for 5 min with PBS and then incubated without light for 1hr at room temperature with Alexa Fluor® 568 Goat Anti-Mouse IgG (H+L) (Invitrogen) secondary antibodies, diluted 1:2000 in IF buffer. After washing 3 times with PBS, 1 x Hoechst 33258 was used for staining DNA.

Chapter III: New iteradensoviruses

Preamble

The *Densovirinae* subfamily contain several genera, the *Ambidensovirus* genus (ambisense densoviruses, a separate NS3 and ITRs), the *Brevidensovirus* genus (monosense viruses with a small 4 kb genome and lacking PLA2 activity), the *Penstyldensovirus* genus (also small, monosense densoviruses infecting shrimp and lacking PLA2 activity) and the *Iteradensovirus* genus with a 5 kb genome with ITRs, a monosense genome organization and virions with PLA2 activity. Until a few years ago, only three iteradensoviruses were known: BmDNV-1 from *Bombyx mori* (Li *et al.*, 2001), CeDNV from *Casphalia extranea* (Fédière *et al.*, 2002), and DpDNV from *Dendrolimus punctatus* (Wang *et al.*, 2005). Over the last three years, we got three different infected insect samples for pathogen detection. These samples including some old, stored *Sibine fusca* larvae infected by a densovirus, *Papilio polyxene* (black swallowtail butterfly) larvae from Professor Ann E. Hajak of Cornell University, and the third sample is some infected pupae of monarch butterfly from a farm in Granby (Quebec). We used a ‘Sequence Independent Single Primer Amplification (SISPA)’ method to identify the unknown pathogens from these insect samples. The detection results revealed three new iteradensoviruses of these insects. These new viruses were later named as PpDNV for *Papilio polyxenes* larvae (Yu *et al.*, 2012b), SfDNV for *Sibine fusca* larvae (Yu *et al.*, 2012a) and DppIDV for monarch butterfly pupae (Yu *et al.*, 2014). All of these new discovered viruses have highly conserved genome organization with previously indentified iteradensoviruses: CeDNV, BmDNV-1 and DpDNV.

Article 1: Iteravirus-Like Genome Organization of a Densovirus from *Sibine fusca* Stoll.

Qian Yu ^a, Gilles Fédière ^a, Adly Abd-Alla ^b, Max Bergoin ^a and Peter Tijssen ^a

INRS-Institut Armand-Frappier, Laval, Quebec, Canada, ^a

Insect Pest Control Laboratory, Joint FAO/IAEA Division, International Atomic Energy Agency, Vienna, Austria ^b

Contribution of authors

Gilles Fédière and Adly Abd-Alla were responsible for initiated isolation of the infected larvae and started the project. Qian Yu was responsible for all the experiments of this work including virus purification from infected larvae, DNA extraction, cloning, and data analysis. Qian Yu, Max Bergoin and Peter Tijssen prepared the manuscript. Peter Tijssen supervised the project.

Abstract

The complete genome of *Sibine fusca* densovirus was cloned and sequenced. The genome contained 5,012 nucleotides (nt), including inverted terminal repeats (ITRs) of 230 nt with terminal hairpins of 161 nt. Its DNA sequence and monosense organization with 3 open reading frames (ORFs) is typical of the genus *Iteravirus** in the subfamily *Densovirinae* of the *Parvoviridae*.

Résumé

Le génome complet du virus *Sibine fusca* densovirus a été cloné et séquencé. Le génome contient 5012 nucléotides (nt), incluant les répétitions terminales inversées (ITRs) de 230nt avec des épingles à cheveux de 161nt. La séquence d'ADN et l'organisation monosens avec 3 cadres de lecture ouverts (ORFs) est typique du genre *Iteravirus** dans la sous famille *Densovirinae* de la famille *Parvoviridae*.

* Note: Since the publication of this manuscript, the genus name have been changed to *Iteradenvirus* (Cotmore et al., 2014a).

Results

The slug caterpillar *Sibine fusca* Stoll (syn. *Acharia fusca*; *Limacodidae*), a major pest of oil palm, is widely distributed in Southern America. The larvae live gregariously in colonies of up to 60 individuals and disperse just before pupation. Virus-infected larvae showed nuclear lesions typical of a densovirus infection (Fédière, 2000, Meynardier et al., 1977). Electron microscopy revealed isometric particles with about 20-nm diameters, also characteristic of densoviruses (Figure 3.1). Infected larvae of *S. fusca* drifted from the feeding colony, and considerable proliferation of cells, resembling tumors, was observed in the midgut (Meynardier et al., 1977). Thus far, this virus has not been further characterized, cloned, or sequenced but has been used effectively in biological control (Genty et al., 1975).

The virus was partially purified by the method described for *Galleria mellonella* densovirus (GmDNV) (Longworth et al., 1968) from an infected larva. A sequence-independent, single-primer amplification (SISPA) method (Reyes et al., 1991) was used in a preliminary genome characterization. DNA, extracted under conditions of high ionic strength to anneal the single-stranded DNA (ssDNA), had a size of around 5 kb. This DNA was digested with the *Csp6I* restriction enzyme, ligated with an adaptor, amplified by PCR as described elsewhere (Allander et al., 2001), and cloned into the PCR2.1 vector by the TA cloning method (Holton et al., 1991). Amplicon inserts were sequenced by Sanger's method as described previously (Tijssen et al., 2003). A unique *Cla*I restriction site was observed near the middle of a preliminary 4.7-kb sequence. DNA from the virus was then blunt-ended by a mixture of Klenow fragment and T4 DNA polymerase, digested with *Cla*I, and cloned into *EcoRV* and *Cla*I sites in the pBluescriptSK(-) vector, yielding clones with a 2.6-kb insert and clones with a 2.4-kb insert. Four inserts of each set were sequenced in both directions using Sanger's

method and the primer-walking method as described before (Tijssen *et al.*, 2003). Insert sequences were identical in each set except for the flip-flop sequences in the hairpins.

The *Sibine fusca* densovirus (SfDNV) genome contained inverted terminal repeats (ITRs) typical of the three members (*Bombyx mori* densovirus type 1 [BmDNV-1], *Casphalia extranea* densovirus [CeDNV], and *Dendrolimus punctatus* densovirus [DpDNV]) of the *Iteravirus* genus and with a length of 230 nucleotides (nt) (Tijssen *et al.*, 2011). The terminal J-shaped hairpins of 161 nt (Figure 3.1) were about 90% conserved between BmDNV-1 (Li *et al.*, 2001), CeDNV (Fédière *et al.*, 2002), and DpDNV (Wang *et al.*, 2005). In the hairpins, nt 60 to 102 and nt 4911 to 4953 occurred in two orientations, “flip” and its reverse complement orientation “flop,” that were identical to the flip-flop of CeDNV and 98% identical to that of BmDNV. The overall sequence was about 85% identical to CeDNV, about 78% identical to BmDNV, and about 72% identical to DpDNV.

The monosense genome contained three intronless genes that were virtually identical in position and size to those of other iteraviruses. The largest open reading frame (ORF), ORF1 (nt 354 to 2615), had a coding capacity of 753 amino acids (aa) and the typical NTPase motif for NS1 (Fédière *et al.*, 2002). ORF2 (nt 2669 to 4714), with the phospholipase A2 motif characteristic for VP (Zádori *et al.*, 2001), had a coding capacity of 681 aa. ORF3 corresponded to NS2 with a 452-aa coding capacity and typically overlapped the N terminus of NS1 (nt 481 to 1839). As a comparison, for the other iteraviruses, the NS1 is 753 to 775 aa, the NS2 is 451 to 453 aa, and the VP is 668 to 678 aa.

Nucleotide sequence accession number. The GenBank accession number of SfDNV is JX020762.

ACKNOWLEDGMENTS

This work was supported by a grant from the Natural Sciences and Engineering Research Council of Canada to P.T. Q.Y. acknowledges support from a scholarship from the People's Republic of China. G.F. was supported by IRD during his sabbatical at INRS. A. A.-A. is supported by IEAE. G.F. and M.B. are invited professors at INRS.

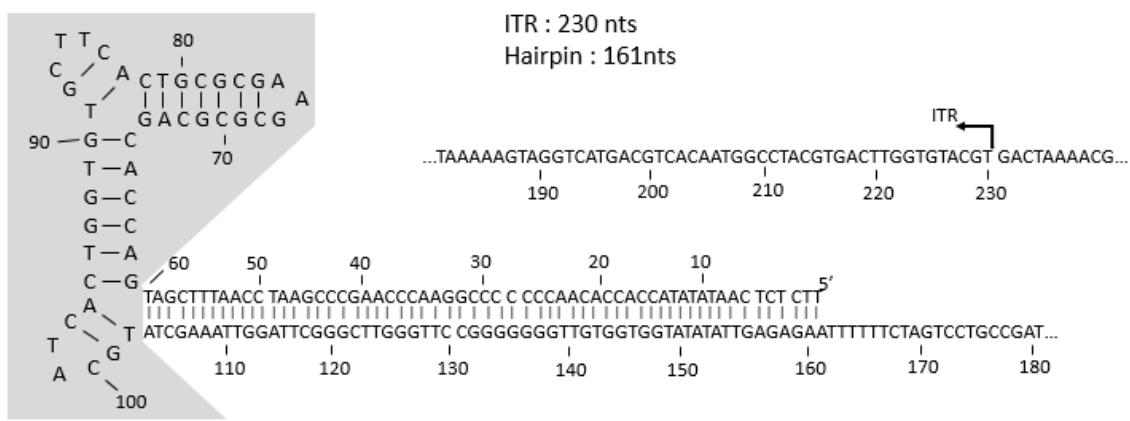
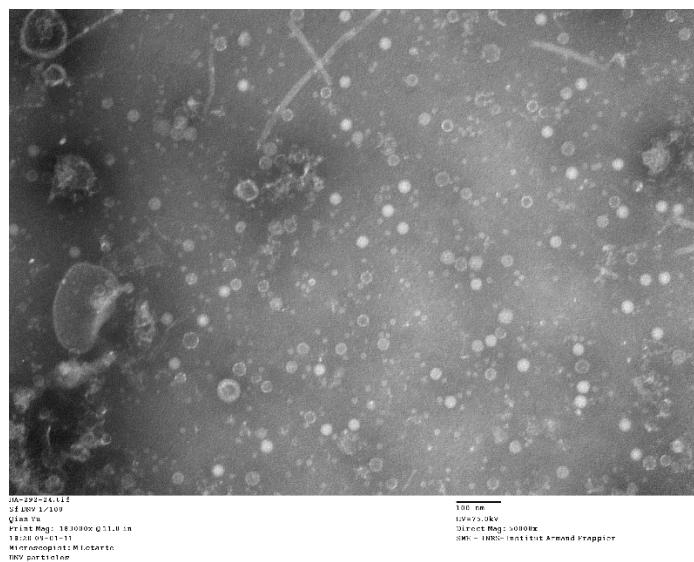


Figure 3.1 EM of SfDV virus particle and J-shaped telomere sequence of SfDV ITR (Scale bar: 100nm).

Article 2: *Papilio polyxenes* Densovirus Has an Iteravirus-Like Genome Organization

Qian Yu ^a, Ann E. Hajek ^b, Max Bergoin ^a and Peter Tijssen ^a

INRS-Institut Armand-Frappier, Laval, Quebec, Canada, ^a

Cornell University, Department of Entomology, Ithaca, New York, USA ^b

Contribution of authors

Ann E. Hajek provided the infected larvae samples. Qian Yu was responsible for conducting all the experiments, analyzing data, and co-responsible for planning. Peter Tijssen and Qian Yu prepared the manuscript. Peter Tijssen supervised the project.

Abstract

The genome of *Papilio polyxenes* densovirus was cloned and sequenced and contained 5,053 nucleotides (nt), including inverted terminal repeats (ITRs) of 271 nt with terminal hairpins of 175 nt. Its DNA sequence and monosense organization with 3 open reading frames (ORFs) are typical of the genus *Iteravirus** in the subfamily *Densovirinae* of the *Parvoviridae*.

Résumé

Le génome du virus *Papilio polyxenes* densovirus a été cloné et séquencé et contient 5053 nucléotides (nt), incluant les répétitions terminales inversées (ITRs) de 271nt avec des épingle à cheveux de 175nt. La séquence d'ADN et l'organisation monosens avec 3 cadres de lecture ouverts (ORFs) est typique du genre *Iteravirus** dans la sous famille *Densovirinae* de la famille *Parvoviridae*.

* Note: Since the publication of this manuscript, the genus name have been changed to *Iteradensovirus* (Cotmore et al., 2014a).

Results

The larvae of the black swallowtail (*Papilio polyxenes*; family *Papilionidae*), a butterfly found throughout eastern North America, feed gregariously on host plants of the carrot family (*Umbelliferae*), such as dill, parsley, and fennel. These plants produce stimulatory compounds for chemoreception of this insect but also furanocoumarins that are toxic and serve as a defense mechanism against various insect predators. *Papilio polyxenes*, which adapted to furanocoumarin-containing host plants, therefore provides an excellent laboratory model to study insect-plant coevolution (Berenbaum *et al.*, 1981, Wen *et al.*, 2006). During recent years, significant mortality was observed in laboratory animals and electron microscopy examination revealed isometric particles of about 20 nm in diameter (F. Pringle, unpublished observations), characteristic of densovirus (Figure 3.2). A significant proportion of the larvae obtained from the field were also infected.

Papilio polyxenes densovirus (PpDNV) was partially purified from 0.5 g larvae as described previously for *Galleria mellonella* DNV (GmDNV) (Longworth *et al.*, 1968) and *Sibine fusca* Stoll DNV (SfDNV) (Yu *et al.*, 2012a). PpDNV DNA, extracted under conditions of high ionic strength to anneal the single-stranded DNA (ssDNA), had a size of around 5 kb. A sequence-independent, single-primer amplification (SISPA) method (Reyes *et al.*, 1991) was used as previously described (Yu *et al.*, 2012a). The amplicons were cloned into the pGEM-T vector by the TA cloning method (Holton *et al.*, 1991) and sequenced by Sanger's method as described previously (Tijssen *et al.*, 2003). A unique *SacI* restriction site detected near the middle of a 4.7-kb sequence was used to clone the blunt-ended PpDNV DNA, obtained from virus, restricted with *SacI* in the *EcoRV* and *SacI* sites of the PCR2.1 vector. Clones with a 1.6-kb insert and clones with a 3.4-kb insert were obtained, and four inserts of each set were sequenced in both directions using Sanger's method and the primer-walking method as described previously (Tijssen *et al.*, 2003). Insert sequences were identical in each set except for the flip-flop sequences.

The overall sequence had a high identity with iteraviruses (identity of about 78% to *Casphalia extranea* DNV [CeDNV], about 75% to *Bombyx mori* DNV [BmDNV-1], about 74% to SfDNV, and about 67% to *Dendrolimus punctatus* DNV [DpDNV]). The PpDNV genome contained typical inverted terminal repeats (ITRs) of the four members of the *Iteravirus* genus (BmDNV-1, CeDNV, SfDNV, and DpDNV), albeit a bit longer (271 versus 230 nucleotides [nt]) (Tijssen *et al.*, 2011). The terminal J-shaped hairpins of 175 nt were about 80% conserved with BmDNV-1 (Li *et al.*, 2001), CeDNV (Fédière *et al.*, 2002), SfDNV (Yu *et al.*, 2012a), and DpDNV (Wang *et al.*, 2005). In the hairpins, nt 67 to 109 and nt 4945 to 4987 occurred in two orientations, “flip” and its reverse-complement orientation “flop,” that were close to 100% identical to the flip-flop of the other iteraviruses (Figure 3.2).

Similar to other iteraviruses, the PpDNV monosense genome contained three intronless genes with essentially identical positions and sizes. The largest, open reading frame 1 (ORF1) (nt 349 to 2631), had a coding capacity of 760 amino acids (aa) and the typical nucleoside triphosphatase (NTPase) motif for NS1 (Fédière *et al.*, 2002). ORF2 (nt 2686 to 4707) with the phospholipase A2 motif typical for parvovirus VP (Zádori *et al.*, 2001) had a coding capacity of 673 aa. ORF3 corresponded to NS2 with a 455-aa coding capacity and typically overlapped the N terminus of NS1 (nt 482 to 1849). As a comparison, NS1 is 753 to 775 aa, NS2 is 451 to 453 aa, and VP is 668 to 678 aa for the other iteraviruses.

Nucleotide sequence accession number. The GenBank accession number for PpDNV is JX110122.

ACKNOWLEDGMENTS

This work was supported by a grant from the Natural Sciences and Engineering Research Council of Canada to P.T. Q.Y. acknowledges support from a scholarship from the People’s Republic of China. M.B. is an invited professor at INRS.

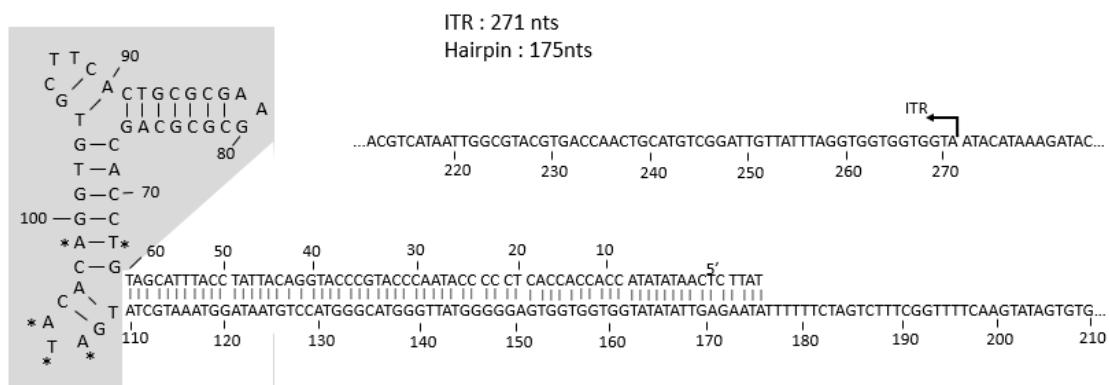
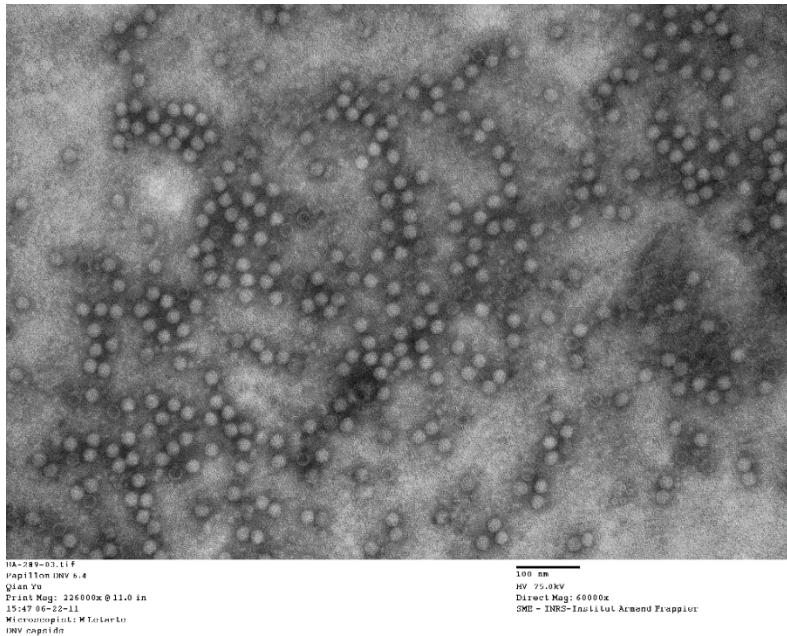


Figure 3.2 EM of PpDV virus particle and J-shaped telomere sequence of PpDV (Scale bar: 100nm).

Article 3: Iteradensovirus from the Monarch Butterfly, *Danaus plexippus plexippus*

Qian Yu and Peter Tijssen

INRS-Institut Armand-Frappier, Université du Québec, 531 boul. des Prairies, Laval, Québec, Canada, H7V1B7

Contribution of authors

Qian Yu was responsible for conducting all the experiments, analyzing data, and co-responsible for planning. Peter Tijssen and Qian Yu prepared the manuscript. Peter Tijssen supervised the project.

Abstract

The 5,006-nucleotide (nt)-long genome of a new virus from monarch butterfly pupae was cloned and sequenced. It was flanked by inverted terminal repeats (ITRs) of 239 nt with 163-nt hairpins. The monosense genome with three open reading frames is typical of the genus *Iteradensovirus* in the subfamily *Densovirinae* of the family *Parvoviridae*.

Résumé

Le génome de 5006 nucléotide (nt) d'un nouveau virus provenant de pupes de papillons monarque a été cloné et séquencé. Il est flanqué de répétitions terminales inversées (ITRs) de 239nt avec des épingle à cheveux de 163nt. L'organisation monosens du génome avec 3 cadres de lecture ouverts (ORFs) est typique du genre *Iteradensovirus* dans la sous famille *Densovirinae* de la famille *Parvoviridae*.

Results

Monarch butterflies (*Danaus plexippus plexippus*) migrate from eastern and central North America for overwintering in Mexico. Migration of this emblematic butterfly has been in rapid decline in recent years, prompting the presidents of the United States and Mexico and the Prime Minister of Canada to discuss this problem during a meeting in February 2014. Several factors may be responsible for this trend. The cool and relatively moist high mountain habitats of Oyamel fir forests are ideal for both the firs and the butterflies. The forest canopy and the clustering of the monarchs protect them against freezing (Anderson *et al.*, 1996). Severe logging and climate change threaten these forests, and a massive reforestation effort is under way to reverse this trend. Second, the extensive use of genetically modified herbicide-resistant soybeans and corn may be reducing the number of larval host plants, milkweeds, especially in their main habitat in the Corn Belt (Brower *et al.*, 2006, Knight *et al.*, 2009, Malcolm *et al.*, 1989), encouraging the suggestion of a milkweed corridor. However, this has been disputed elsewhere (Niiler, 1999). Third, pathogens such as bacteria, parasites, and viruses may affect monarch populations (Arnott *et al.*, 1968, Bartel *et al.*, 2011, De Roode *et al.*, 2011).

Virus was partially purified from three infected pupae obtained from a butterfly farm in Granby (Quebec, Canada) by the method described for *Galleria mellonella* densovirus (Tijssen *et al.*, 2003) and visualized by electron microscopy (Figure 3.3). A preliminary genome characterization was obtained with the sequence-independent single-primer amplification (SISPA) method (Allander *et al.*, 2001, Yu *et al.*, 2012a, Yu *et al.*, 2012b), showing two *Spel* restriction sites in a preliminary 4.7-kb sequence. Viral DNA was then blunt ended by a mixture of Klenow large-fragment and T4 DNA polymerase, digested with *Spel*, and cloned into *EcoRV* and *Spel* sites in the pBluescriptSK II(-)vector, yielding clones with 3.4-kb inserts and clones with 1.5-kb inserts. Sequences of several complete clones, obtained in both directions with Sanger's method (Yu *et al.*, 2012a, Yu *et al.*, 2012b) were identical except for the flip-

flop sequences in the hairpins. The sequence between the two *Spel* sites was obtained after PCR amplification with gene-specific primers.

The *D. plexippus plexippus* iteradensovirus (DppIDV) genome contained the typical inverted terminal repeats (ITRs) of members of the *Iteradensovirus* genus (*Bombyx mori* densovirus 1 [BmDNV-1], *Casphalia extranea* densovirus [CeDNV], *Sibine fusca* Densovirus [SfDNV], *Papilio polyxenes* densovirus [PpDNV], and *Dendrolimus punctatus* densovirus [DpDNV]) (Fédière *et al.*, 2002, Li *et al.*, 2001, Wang *et al.*, 2005, Yu *et al.*, 2012a, Yu *et al.*, 2012b). The 239-nucleotide (nt) ITRs with 163-nt terminal J-shaped hairpins were about 90% conserved with those of the other iteradensoviruses (Figure 3.3). The overall sequence was about 86% identical to CeDNV, about 84% identical to SfDNV and BmDNV, about 78% identical to PpDNV, and about 71% identical to DpDNV.

Similar to other iteradensoviruses, the DppIDV monosense genome contained three intronless genes with essentially identical positions and sizes. The largest, open reading frame 1 (ORF1) (nt 360 to 2618), had a coding capacity of 752 amino acids (aa) and the typical nucleoside triphosphatase (NTPase) motif for NS1. ORF2 (nt 2677 to 4710), with the phospholipase A2 motif, typical for parvovirus VP, had a coding capacity of 677 aa. ORF3, with a 451-aa coding capacity (nt 487 to 1842) corresponded to NS2 and overlapped NS1 at its N terminus. As a comparison, NS1 is aa 753 to 775, NS2 is aa 451 to 455, and VP is aa 668 to 681 for the other iteradensoviruses.

Nucleotide sequence accession number. The GenBank accession no. for DppIDV is KF963252.

ACKNOWLEDGMENTS

This work was supported by the Natural Sciences and Engineering Research Council of Canada grant to P.T. Q.Y. acknowledges support from a scholarship from the People's Republic of China and tuition waivers from INRS-Institut Armand-Frappier.

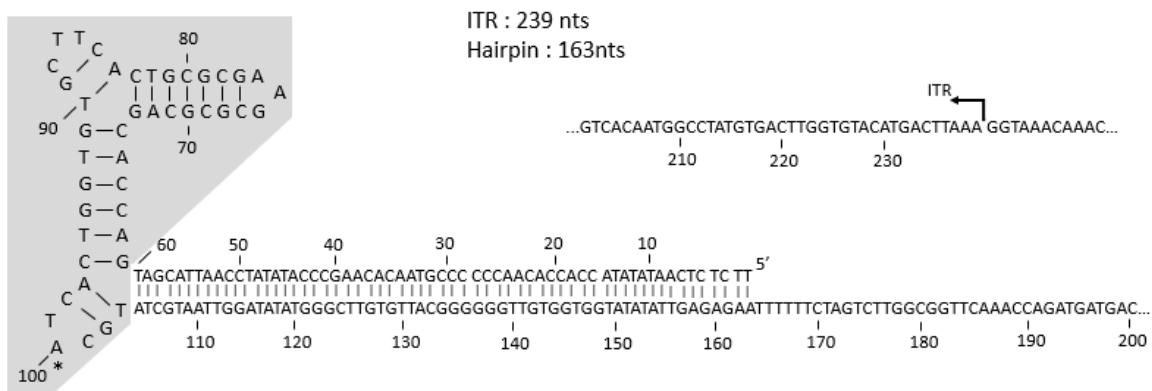
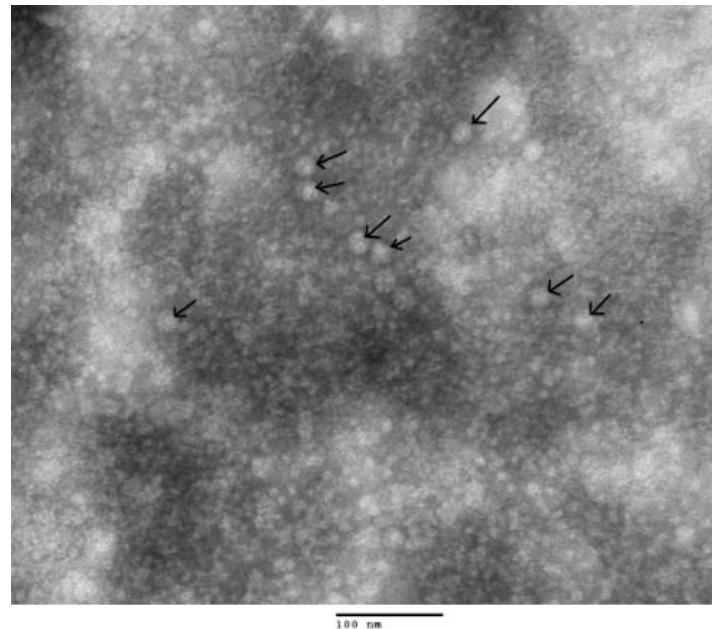


Figure 3.3 EM of DppIDV virus particle and J-shaped telomere sequence of DppIDV (Scale bar: 100nm).

Chapter IV: Iteradensoviruses transcription strategies

Preamble

None of the expression strategies of iteradensoviruses have been reported, due to the lack of a suitable cell line to support viral genome replication. Thus, my project about iteradensoviruses then focused on determining the transcription strategy of the *Iteradensovirus* genus. We first obtained the infectious clones of the three new iteradensoviruses. Together with the infectious clone of CeDNV and BmDNV-1, we tried to produce mRNA of those five viruses in different insect cell lines. Finally, the LD652 cell line was selected for the transcription study. For PpDNV, we got also the host insect *Papilio polyxenes* larvae from the Granby farm to support the viral mRNA production. Using the classical method for identifying the 5' and 3' ends of NS and VP transcripts, we got the complete transcription maps for the five iteradensoviruses. Finally, our results suggested an expression strategy for all members of the *Iteradensovirus* genus.

Article 4: Gene expression of five different iteradenoviruses: BmDV, CeDV, PpDV, SfDV and DpIDV

Qian Yu and Peter Tijssen

INRS-Institut Armand-Frappier, Université du Québec, 531 boul. des Prairies, Laval, Québec, Canada, H7V1B7

Contribution of authors

Qian Yu was responsible for conducting all the experiments, analyzing data, and co-responsible for planning. Peter Tijssen and Qian Yu prepared the manuscript. Peter Tijssen supervised the project.

Abstract

Iteradenoviruses are 5 kb parvoviruses with typical J-shaped inverted-terminal repeats of about 250 nucleotides and terminal hairpins of about 165 nucleotides. The single-stranded DNA genome contains several open reading frames but their expression strategy is still unknown. Here the transcription maps and expression of the viruses in this genus were explored. As for brevidenoviruses, the two NS genes were expressed by overlapping promoters with alternate transcription starts at both sides of the NS1 start codon.

Résumé

Les Iteradenovirus sont des parvovirus avec un génome de 5kb avec des répétitions terminales inversées en forme de J d'environ 250 nucléotides ainsi que des épingle à cheveux d'environ 165 nucléotides. Le génome d'ADN simple brin contient plusieurs cadres de lecture ouverts mais leur stratégie d'expression demeure inconnue. Ici nous avons caractérisé les stratégies d'expression de virus de ce genre. Tout comme les brevidenovirus, les deux gènes NS sont exprimé avec des promoteurs chevauchants avec des codons d'initiation alternatifs de part et d'autre du codon d'initiation de NS1.

Introduction

Invertebrate densoviruses (DVs) form a separate subfamily (*Densovirinae*) within the *Parvoviridae* family (Cotmore *et al.*, 2014a, Gudenkauf *et al.*, 2014, Kapelinskaya *et al.*, 2011, Tijssen *et al.*, 2011) and have recently been reclassified (Cotmore *et al.*, 2014a). They share many physicochemical and genome properties with *Parvovirinae* of vertebrates (Chapman *et al.*, 2006, Cotmore *et al.*, 2006a, Tijssen *et al.*, 2006a) and most viruses of these subfamilies have a phospholipase A2 (PLA2) activity for cell entry (Canaan *et al.*, 2004, Farr *et al.*, 2005, Zádori *et al.*, 2001). The left-hand side of the genome contains the non-structural (NS) genes with the rolling-circle replication endonuclease and the superfamily 3 helicase in NS1 (Cotmore *et al.*, 2006b, Nüesch, 2006). Open-reading frames (ORFs) for structural proteins (VPs) are on the right-hand side of the genomes (Cotmore *et al.*, 2006a, Tijssen *et al.*, 2006a).

Densovirinae contain *Ambidensovirus*, *Brevidensovirus*, *Hepandensovirus*, *Penstyldensovirus* and the *Iteradensovirus* genera which have distinct genome sizes and organizations, and presence or absence of inverted terminal repeats (ITRs) and PLA2 (Cotmore *et al.*, 2014a, Tijssen *et al.*, 2011). Iteradensoviruses have a 5 kb genome with ITRs, a monosense genome organization and PLA2 activity. Four previously-known iteradensoviruses are BmDV from *Bombyx mori* (Nakagaki *et al.*, 1980), CeDV from *Casphalia extranea* (Fédière *et al.*, 2002), DpDV from *Dendrolimus punctatus* (Wang *et al.*, 2005) and HaDV2 from *Helicoverpa armigera* (Xu *et al.*, 2012). The reported sequence of HaDV2 is incomplete, differs significantly from the others and remains to be confirmed. Recently we isolated PpDV from *Papilio polyxenes* (black swallowtail butterfly) (Yu *et al.*, 2012b), SfDV from *Sibine fusca* (oil palm pest) (Yu *et al.*, 2012a) and DpIDV from Monarch butterfly (Yu *et al.*, 2014) (Figure 4.1A). Based on NS1 identities, taxons of these viruses have been defined as Lepidopteran Iteradensovirus LI 1 with BmDV, the LI 2 species with CeDV, SfDV and DpIDV, species LI 3 with DpDV, species LI 4 with PpDV and species LI 5 with HaDV2 (Cotmore *et al.*, 2014a). NS1 and NS2 proteins of the Chinese isolates (LI 3 and 5) had <40% identity

scores with other iteradensoviruses. In contrast, except for HaDV2, the VP proteins of iteradensoviruses have all similar identities of 70-80%, independent of species.

Infectious clones were created and sequenced for BmDV (pIN919), CeDV (pSMART-CeDV), PpDV (pCR2.1-PpDV), SfDV (pBlue-SfDV) and DpIDV (pBlue-DpIDV) (Fédière *et al.*, 2002, Li *et al.*, 2001, Yu *et al.*, 2012a, Yu *et al.*, 2012b, Yu *et al.*, 2014). The completely sequenced genomes have all typical J-shaped ITRs with lengths ranging from 230-271 nucleotides (nts) and terminal hairpins of 159-175 nts (Figure 4.1B, C). The 5'- and 3'-terminal hairpins occurred in two orientations, “flip” and its reverse-complement orientation “flop.” These were identical among CeDV, SfDV and DpIDV and had few differences in PpDV and BmDV. Positions and sizes of ORFs were virtual identical in all iteradensoviruses (Figure 4.2). The typical parvoviral phospholipase A2 motif (Zadori *et al.*, 2001) was present in all VP1s.

Densovirus transcription has not been studied extensively. So far, VP transcripts of different densoviruses with one VP ORF are not spliced but sets of N-extended proteins are generated by leaky scanning (Fédière *et al.*, 2004, Pham *et al.*, 2013a, Tijssen *et al.*, 2006a, Tijssen *et al.*, 2003). Among others, BgDNV and AdDNV have two VP ORFs that are often joined by splicing and generate proteins with alternate N-termini in addition to nested N-terminally extended sets by leaky scanning (Kapelinskaya *et al.*, 2011, Liu *et al.*, 2011). Leaky scanning is also important for the NS proteins of densoviruses from different genera (Fédière *et al.*, 2004, Kapelinskaya *et al.*, 2011, Liu *et al.*, 2011, Pham *et al.*, 2013a, Tijssen *et al.*, 2006a, Tijssen *et al.*, 2003). However, splicing determines whether NS3 or NS1/NS2 is expressed for the ambisense densoviruses whereas the two NS genes of the brevidensoviruses have overlapping promoters so that transcripts start at either side of the NS1 starting codon allowing either NS1 or the downstream NS2 to be expressed (Fédière *et al.*, 2004, Liu *et al.*, 2011, Pham *et al.*, 2013a, Tijssen *et al.*, 2003). So far transcription of viruses in the *Iteradensovirus* genus has not been studied.

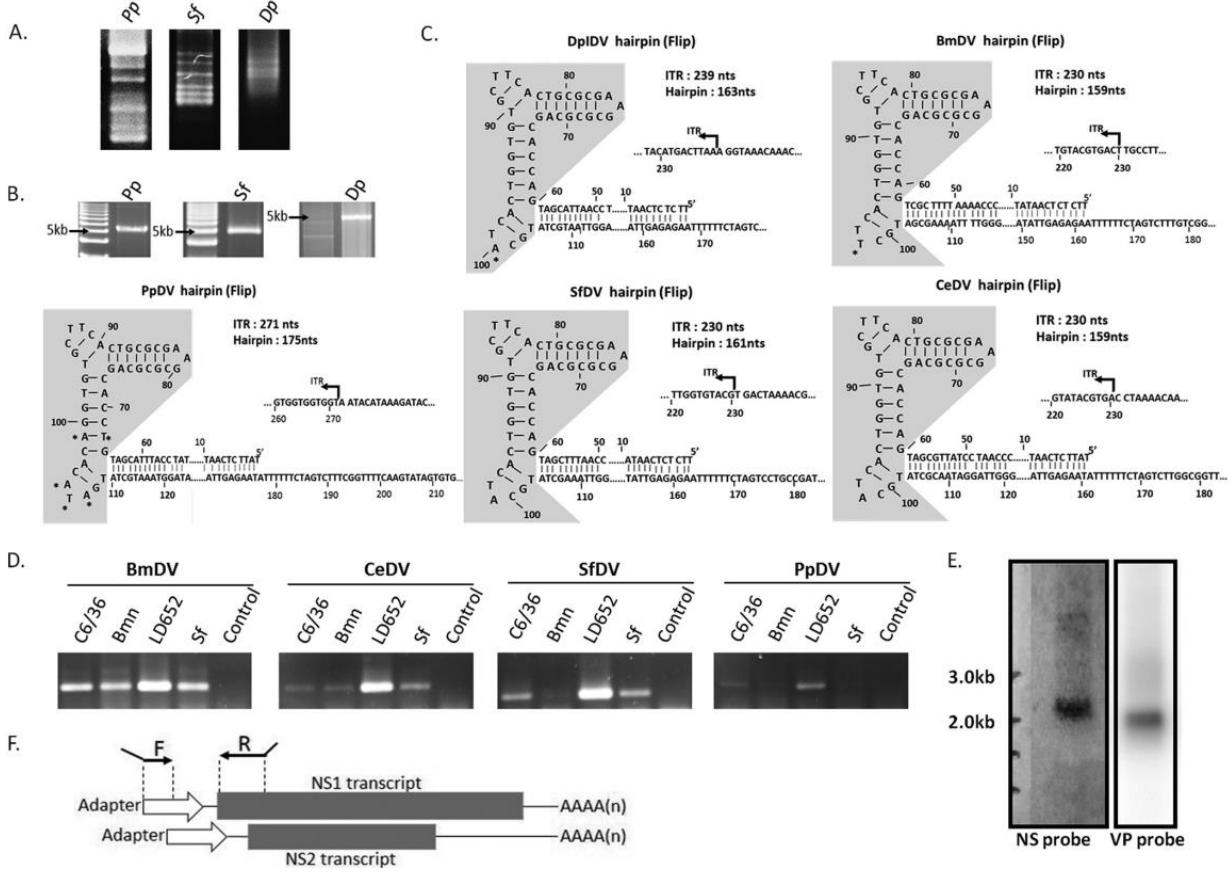


Figure 4.1 Iteradensoviruses and their transcription.

(A) The detection of pathogens by Sequence-independent single-primer amplification in larvae from *Papilio polyxenes* (Pp) and *Sibine fusca* (Sf) and in pupae from the Monarch butterfly (Dp) on a 2% agarose gel. (B) DNA purification of new iteradensoviruses on 1% electrophoresis agarose gels. DNA ladder: 1kb plus DNA ladder. (C) Inverted terminal repeats (ITRs) of 5 iteradensoviruses with J-shaped terminal hairpins. The shaded areas in the ITRs occur in two alternative sequences, named “flip” and its reverse-complement “flop”. Iteradensoviruses’ ITRs sequences are highly similar to each other. Only a few nucleotides differ and were indicated with an asterisk. (D) Infectious clones of CeDV, BmDV, PpDV and SfDV were transfected into 4 different insect cell lines (C6/36, Bmn, Ld652 and Sf) and total RNA was extracted 48hr post-transfection. Gene specific primers were used for RT-PCR identification. PCR products were checked by electrophoresis in 1% agarose gel. Total RNAs without reverse transcription were used as a control template for PCR. (E) Northern blotting of PpDV NS and VP transcripts. Both NS and VP yielded a single transcript bands. However, further results showed that the 2.3kb transcripts for NS consisted of two transcripts with almost identical size. (F) Scheme of a PCR method for distinguishing NS1 and NS2 transcription start. A forward primer with ~20 nts complementary sequence with the RACE adapter at 3' and an unspecific extension with restriction sites at the 5'-end. The 3'-end of the reverse primers were extended with the sequence complementary to the putative NS1 ATG.

Results

Production of iteradensovirus mRNAs and Northern blotting.

A major impediment to iteradensovirus studies has been the lack of suitable insect cells that support their replication. Four different insect cell lines were compared for their efficacy to produce iteradensovirus mRNA (Figure 4.1D). LD652, Sf and Bmn cells (in Sf900(II) medium) and mosquito C6/36 cells (in RPMI 1640 medium) were propagated with 5% and 10% fetal bovine serum, respectively. Transfections with the infectious clones of CeDV, BmDV, PpDV and SfDV were done using DOTAP Liposomal Transfection Reagent (Roche). PpDV was purified as previously described (Tijssen *et al.*, 1977). Third-instar *Papilio* larvae, the only host insect available, were fed pesticide-free parsley leaves that had been submerged in a PpDV suspension and served also as a source of PpDV mRNA. RT-PCR results (Figure 4.1D) showed that LD652 cells were most suitable for mRNA production of these iteradensoviruses. Total RNA was isolated from *Papilio* larvae 72h post-infection and 48 h post-transfection from LD652 cells transfected with SfDV, BmDV, CeDV or DpIDV infectious clones by the NucleoSpin RNA II Kit (Clontech). An extra DNaseI treatment was added after RNA extraction with the Turbo DNA-free kit (Ambion). An RT-negative PCR test was included to verify the absence of DNA. Total RNA was subjected to an mRNA purification using an Poly(A)Purist™ Kit (Ambion). Northern blots were obtained as described previously (Pham *et al.*, 2013a). RNA probes for NS and VP gene were transcribed from NS and VP amplicons obtained by PCR using primers PpNSRpF700 and PpNSNpR850 for NS gene, PpVPRpF3200 and PpVPNpR3400 for VP gene (Table 4.1). Northern-blotting of PpDV mRNA revealed one 2.3kb band for the NS gene, and one 2.0kb band for the VP gene (Figure 4.1E). Since the NS2 ORF overlapped the N-terminal half of NS1, it is expected to share the same poly(A) motif in the 3'-ends of their transcripts.

Table 4.1 Primers for the experiments.

Primer name	Sequence 5'→3'	Purpose
Nbam24	AGGCAACTGTGCTATCCGAGGGAG	SISPA adapter and primer
NCsp11	TACTCCCTCGG	SISPA adapter and primer
Pa140	GTT GTG GTG GTA TAT ATT GAG AAT AT	Highly conserved primer in literadenoviruses' ITR
Pp5NS-1	GTCAAGCACTTCCTGAGGTTAA	5'-RACE primers for PpDV NS
Pp3NS2-1	ATAGAACAGAGCCAATACCACATGG	3'-RACE primers for PpDV NS2
Pp3NS1-1	ATGCCCATATGGGGTGCCTACAGGT	3'-RACE primers for PpDV NS1
Pp5VP-1	CTCCTGTACGTGGATCAACTTGTGC	5'-RACE primers for PpDV VP
Pp3VP-1	GTAAATGTCATGACTCCATACCAAGGT	3'-RACE primers for PpDV VP
Sf5NS-1	GTTCCTTCGAAGCTCAATGGAGTC	5'-RACE primers for SfDV NS
Sf3NS2-1	AGCCAGCAATCGAGCCAGTCGGAAG	3'-RACE primers for SfDV NS2
Sf3NS1-1	AACATACAATGCCAATATGGGGTGC	3'-RACE primers for SfDV NS1
Sf5VP-1	TGCTGCTGTATCCATAGGTACTTC	5'-RACE primers for SfDV VP
Sf3VP-1	ATGACTCCAATTCCAGGTTAGAA	3'-RACE primers for SfDV VP
Bm5NS-1	GTTC CATTGTTGATAACTGACAGGG	5'-RACE primers for BmDV NS
Bm3NS1-1	CAGTACTTGATTTATTGGAACACTG	3'-RACE primers for BmDV NS2
Bm3NS2-1	GGCCAACATGGCCCTTGGAACAC	3'-RACE primers for BmDV NS1
Bm5VP-1	CGCTTGCCTGCTAGCTATTAAATC	5'-RACE primers for BmDV VP
Bm3VP-1	GCCTAAGTTATGATTGGATTGTAAAC	3'-RACE primers for BmDV VP
Ce5NS-1	GCTCAATGGAGTCGTGAGAACATT	5'-RACE primers for CeDV NS
Ce3NS1-1	CAATGAATATGACCTACAGCGAAC	3'-RACE primers for CeDV NS2
Ce3NS2-1	CCACATACTCACGCCCTGCCAA	3'-RACE primers for CeDV NS1
Ce5VP-1	CTCCTCCAACATCAGTGGCTCC	5'-RACE primers for CeDV VP
Ce3VP-1	GCTTAACTGCAACTAAATAT	3'-RACE primers for CeDV VP
Dpp5NS-1	GCTGAGAATTCTCGGTTAAGTATT	5'-RACE primers for DpIDV NS
Dpp3NS-1	GGAAGCCAACGTTAGAAGACACCTG	3'-RACE primers for DpIDV NS
Dpp5VP-1	CTCGTCTAGCTGTTCTCGCTCTATTG	5'-RACE primers for DpIDV VP
Dpp3VP-1	CGAAGACAATTCCCTGTTAACTGC	3'-RACE primers for DpIDV VP
PpNSRpF700	GCTTATCAACAAATGGAACATGC	NS probe primer for PpDV
PpNSRpR850 (+T7 sequence)	CAAATTA <u>ATACGACTCACTATAGGCTCGGTATTGATGGAGTCG</u>	NS probe primer for PpDV
PpVPRpF3200	GCTCCCTAGCCTAAATCAACATCATC	VP probe primer for PpDV
PpVPNpR3400 (+T7 sequence)	CAAAT <u>TAATACGACTCACTATAGGGAGAGCTGTGTAATCTCTG</u>	VP probe primer for PpDV
pIZ5-SfNS1FK	CGGGGTACCGTATAAGAGAGCATACTAAACATG	pIZT/V5-SfNS cloning primers
pIZ5-SfNS2RX	GCTCTAGACCAAATGTAAAGGGGCCATG	pIZT/V5-SfNS cloning primers
pIZ5-SfNS2FA	CCACCGGTATACAAAAGATTGGCTAGAAAGTG	pIZT/V5-SfNS cloning primers
pIZ5-SfNS1RAha (+ HA tag)	CCACCGGT <u>AGCGTAATCTGAAACATCGTATGGTAGGAGATATCGATAACATACTTAC</u>	pIZT/V5-SfNS cloning primers
SfNS1mATGFn	GAGCATACTAAACACCTCAAGTCAGTCAGTCGGCAG	pIZT/V5-SfNS mutagenesis primers
SfNS1mATGRn	CTGCCAGACTGACTTGAGGTGTTAATGTATCGTC	pIZT/V5-SfNS mutagenesis primers
SfNS2mATGFn	CTCATTGTTAAACCCCTAATACTACTGGTG	pIZT/V5-SfNS mutagenesis primers
SfNS2mATGRn	CACCAAGTAGTATTAGGGTTATTAACAAATGAG	pIZT/V5-SfNS mutagenesis primers
pIZ5-V5muFn	CCCGCGGTTCGAAGGCAAGCCTATCCAAACCCCTCCTCGG	pIZT/V5-SfNS mutagenesis primers
pIZ5-V5muRn	CCGAGGAGAGGGTTGGGATAGGCTTGCTTCAACCGCGGG	pIZT/V5-SfNS mutagenesis primers

* In the primer names, the "R" and "F" represent reverse and forward primer, respectively. The numbers in the primer names indicate the 5' end of the primer sequence in virus. The T7 and HA-tag sequences were underlined.

Transcript mapping.

Transcription maps for RNA isolated from PpDV-infected *Papilio* larvae or for RNAs isolated from transfected LD652 cells were established by a FirstChoice RLM RACE kit (Ambion) according to the supplier's instructions (primers in Table 4.1). All RACE PCR products, cloned into pGEM-T or pGEM-T easy vectors (Promega), were sequenced by Sanger's method and demonstrated that 5'-ends of NS2 transcripts started at the conserved CAGT sites downstream of the expected NS1 ATG (Figure 4.2A).

Due to the >40-fold NS2 transcript excess, both in cells and larvae, it was challenging to obtain NS1 transcripts for conventional RACE methods and RNase Protection Assays. Therefore, a reverse primer (R in Figure 4.1F) was designed that extended upstream of the NS2 transcript to the ATG of the putative NS1 start to obtain NS1-specific amplicons. Sequencing of these amplicons revealed the 5'-start of the NS1 transcripts (Figure 4.2A). No differences were observed between mRNA from transfected cells or from PpDV-infected larvae. The 5'-ends of NS1 transcripts started usually 2-4 nts upstream of the NS1 ATG (Figure 4.2A). Unexpectedly, 5' untranslated regions of either 4 or 26 nts were obtained by RACE experiments for SfDV NS1 transcripts. NS1 and NS2 3'-ends shared the same poly(A) motif downstream of the TATA box of VP transcripts. The BmDV NS transcript poly(A) was also located downstream of the TATA box of VP transcript, even though this virus contained an intergenic direct repeat of 45 nts between NS1 and VP (Figure 4.2B).

All VP transcripts had short untranslated sequences at 5'-ends, the transcripts started only 10 to 15 nucleotides upstream of the first ATG (Figure 4.2D) that, moreover, had unfavorable Kozak sequences promoting a leaky scanning translation mechanism for VPs as for ambidenviruses (El-Far *et al.*, 2012, Fédière *et al.*, 2004, Pham *et al.*, 2013a, Tijssen *et al.*, 2006a, Tijssen *et al.*, 2003). Previously, we expressed the VP gene of BmDV to generate virus-like particles in order to obtain the 3D structure by X-ray crystallography (Kaufmann *et al.*, 2011). Poly(A) motifs of the 2.0kb VP transcripts

3'-RACE overlapped with stop codons of VP proteins for all iteradensoviruses. It would be interesting to obtain transcript maps for LI3 and 5 viruses as well although the genome organization is identical.

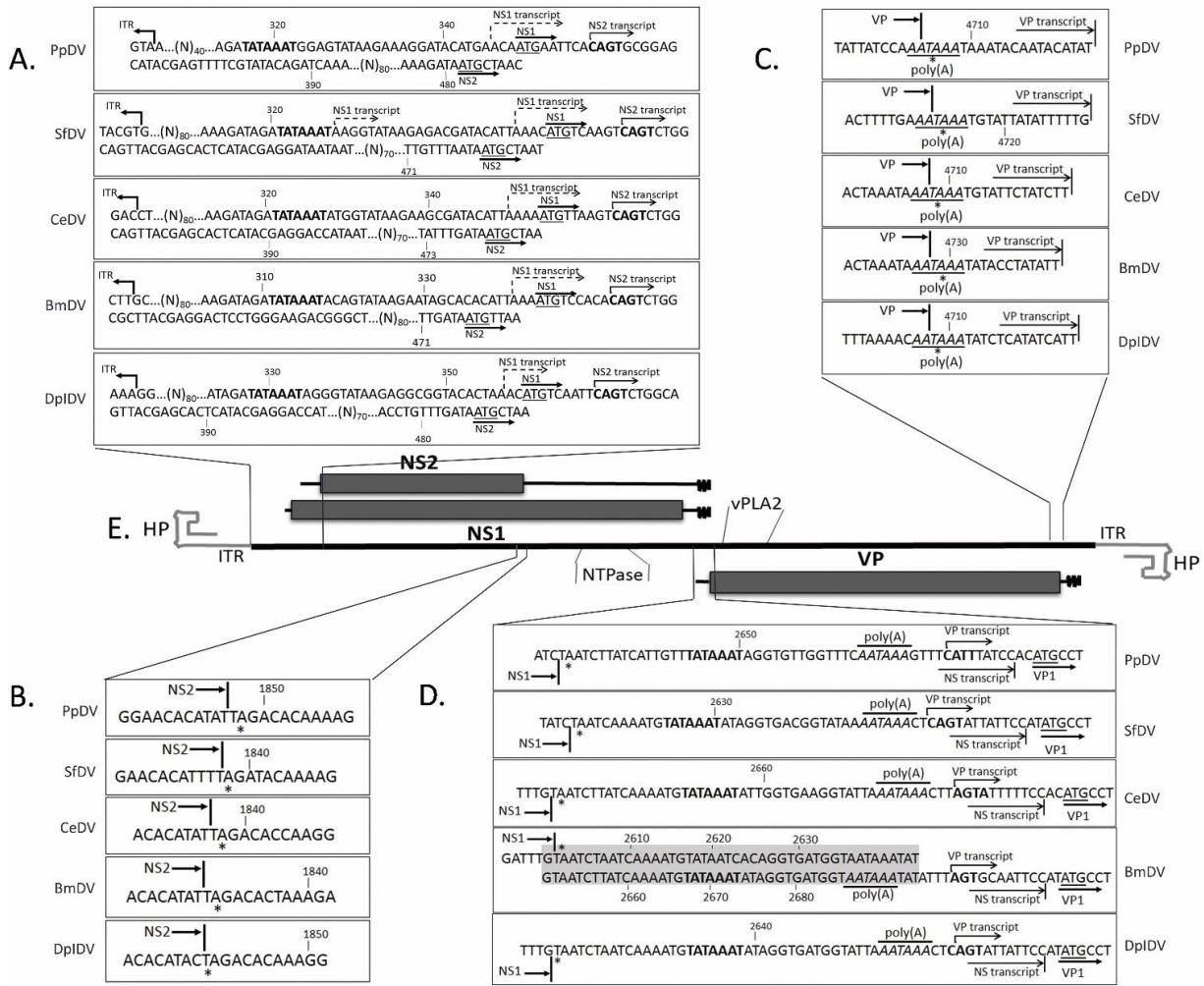


Figure 4.2 Genome organization and transcription profiles of 5 iteradensoviruses.
(A) 5' ends of NS transcripts, the ATGs of NS1 and NS2 ORFs are underlined. CAGT motifs of NS2 transcription starts are indicated in bold. NS1 transcription starts are marked with dashed arrows. **(B)** Stop codons of NS2 are indicated with an asterisk. **(C)** Ends of VP translation and 3' ends of VP transcripts. **(D)** The 3'-ends of NS transcripts and 5-starts of VP transcripts. Poly(A) motifs of NS transcripts are in italics. Translation starts of VP are underlined. **(E)** Genome organization of iteradensovirus genomes is deduced from open reading frames (ORFs) of all 5 iteradensovirus genomes in this report.

Expression of non-structural proteins NS1 and NS2.

The pIZT/V5-His insect cell expression vector was used since it contained an OPIE2 promoter for constitutive expression of NS, the Zeocin resistance gene for colony selection and the Cycle 3 GFP gene for cell line transfection efficiency detection. A V5-epitope was added to the C-terminus of NS2 and a HA-epitope to the C-terminus of NS1 (Figure 4.3A). An amplicon from NS1 ATG to NS2 stop codon with CC added to reverse primer was cloned into the *Kpn*I and *Xba*I cloning sites of the pIZT/V5-His vector to obtain an in-frame NS2 gene/V5-epitope construct. A second amplicon from the NS2 stop codon to NS1 stop codon was obtained by PCR with a HA-tag added to the reverse primer and a mutation in the NS2 stop codon in the forward primer (Table 4.1). ATG of NS1 and NS2 gene was substituted into ACC (primers, Table 4.1). To test SfDV NS expression by their own promoter, the OPIE2 promoter of the vector was replaced by a fragment between *Bsp*H I and *Spe*I sites of the viral genome (nts 193-750; Figure 4.3A). Two additional mutants were created with the TATA-box changed to GAGA and the CAGT motif to TTGT. LD652 cells in 24-wells plates were transfected with 0.5 µg DNA and 5 µl DOTAP reagent. Cell fixation, indirect immunofluorescence (IF) with 1:200 of V5 antibody (Invitrogen) for staining of NS2 or 1:50 of HA antibody (Santa Cruz) for staining of NS1 and DNA staining with Hoechst 33258 were done as described elsewhere (Boisvert *et al.*, 2010).

Compared to positive controls, only NS2 protein expression could be detected when NS1-ATG is mutated whereas after NS2-ATG mutant transfection the NS1 protein expression can be detected (Figure 4.3C). IF results showed that after transfection of LD652 cells with plasmids containing the original presumed viral promoter element, only the NS2 protein expression was detected. After introduction of a stop codon after the NS2 ATG or mutation of the CAGT of NS2 transcript, the NS2 expression became undetectable. Mutation of the TATA box did not change NS2 expression, which may mean that this TATA box may not be an essential element of the NS2 promoter. The viral genome fragment that replaced the OPIE2 promoter may not contain the complete NS1 promoter since NS1 could not be expressed in LD652

cells after transfection. After introducing a HA-tag in frame with NS1 into the SfDV whole genome clone (except the hairpins) (Figure 4.3B), the plasmid expressed NS1 in LD652 cells after transfection. However, NS1 became undetectable after introducing a stop codon after NS1 ATG confirming its authenticity as the NS1 starting codon and its location upstream of the NS2 transcription start. Combined, these results indicated that iteradensovirus have, like brevidensovirus (Pham *et al.*, 2013a), overlapping NS promoters that are responsible for different transcript starts and likely dictate the relative amounts of these transcripts.

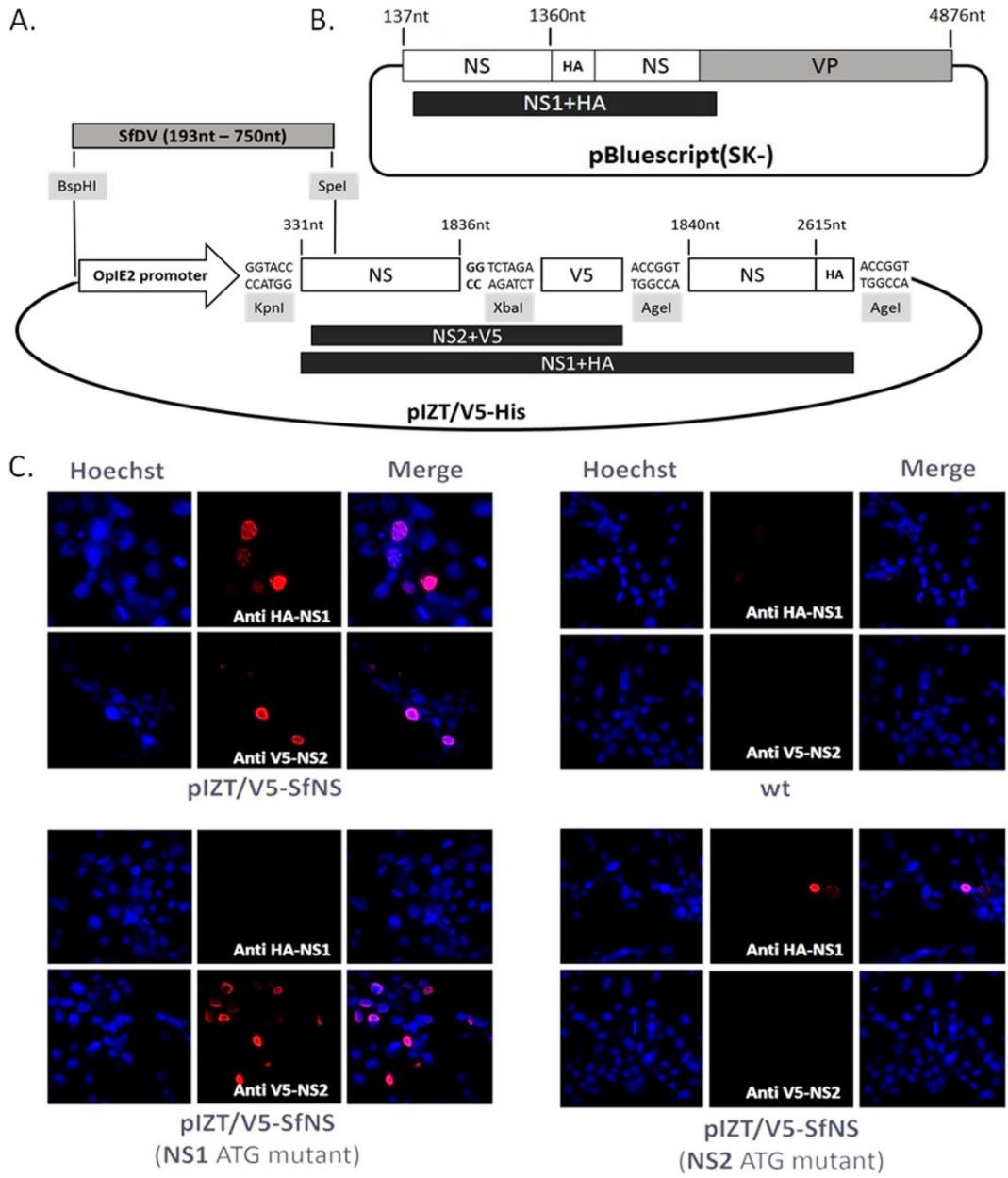


Figure 4.3 Expression of non-structural proteins NS1 and NS2 of SfDV.

(A) SfDV NS gene expression in pIZT/V5 vector. Fragment from 331nt to 1836nt of SfDV genome was first cloned between *KpnI* and *XbaI* site of pIZT/V5 MCS, the “CC” shown in bold was added to have the NS2 in frame with V5 epitope. After the V5 sequence, the 1840nt to 2615nt of SfDV genome was cloned into the *AgeI* site, a HA-tag sequence was added in frame with NS1 to 3’ end of the insert. To investigate the NS promoter element, the OplE2 promoter of vector was replaced in alternate constructs by a fragment from 193nt to 750nt of the SfDV genome, between *BspHI* and *Spel* restriction sites. (B) Scheme for introducing a HA-tag into the SfDV genome clone to in frame with NS1 ORF. SfDV genome sequence from 137nt to 4876nt (lacking the hairpins) was cloned into pBluescript(SK-) vector. HA-tag was added at 1360nt-position of the virus genome. (C) Detection of NS1 and NS2 expression in LD652 cells after 48h post-transfection by immunofluorescence using the HA-tag antibody (for NS1) and V5-antibody (for NS2) and an Alexa Flour 568 Goat-anti-mouse secondary antibody.

ACKNOWLEDGMENTS

This work was supported by a The Natural Sciences and Engineering Research Council of Canada (NSERC) grant to P.T.; Q.Y. acknowledges support from a scholarship from the People's Republic of China and tuition waivers from INRS-Institut Armand-Frappier.

Chapter V: Cricket densoviruses

Preamble

After an initial outbreak of an AdDNV epidemic, many cricket samples were forwarded to our lab for analysis. These analyses were done by my colleague Hanh Thi Pham and myself. We discovered several new cricket viruses in AdDNV-negative samples, these were cloned and sequenced. In particular, there was one new circular ssDNA virus (Vovolvirus), a smaller ambisense densovirus (Mini Ambidensovirus), and very interestingly, a segmented densovirus. Parvovirus with a segmented genome have not been observed thus far. My contribution has been focused on the transcription strategy of AdDNV, a significant part of the work on the Mini Ambidensovirus, and equally on the segmented densovirus (a mosquito densovirus that through recombination changed its tropism to crickets). Finally, we observed a recombination of the AdDNV VP gene with a cricket aphrodisiac iridescent virus.

This chapter combines research on these various cricket viruses that contributed about half of my Ph.D. study.

Article 5: The *Acheta domesticus* densovirus, isolated from the European house cricket, has evolved an expression strategy unique among parvoviruses.

J Virol. 2011 Oct; 85(19):10069-78

Liu K^{1,2}, Li Y^{1,2}, Jousset FX³, Zadori Z¹, Szelei J¹, Yu Q¹, Pham HT¹, Lépine F¹, Bergoin M^{1,3}, Tijssen P¹

1 INRS-Institut Armand Frappier, Université du Québec, Laval, Québec, Canada

2 Current address: Central China Normal University, Wuhan 430079, People's Republic of China

3 Laboratoire de Pathologie Comparée, Université Montpellier 2, Montpellier 34095, France

Contribution of authors

Françoise-Xavière Jousset was responsible for some original AdDNV clones and started the project. Yi Li made the infectious clone and Kaiyu Liu, a postdoc, was the lead investigator. Z. Zadori and J. Szelei did RACE and promoter activity experiments. Qian Yu contributed to express VP clones for transcriptional analysis, purified native viruses. H.T. Pham carried out SDS-PAGE, Western blot and prepared samples for mass spectrometry. F. Lépine was responsible for mass spectrometry analysis. Peter Tijssen and Max Bergoin planned and supervised the project. Peter Tijssen prepared the manuscript. Other authors contributed to review the manuscript before submission.

Abstract

The *Acheta domesticus* densovirus (AdDNV), isolated from crickets, has been endemic in Europe for at least 35 years. Severe epizootics have also been observed in American commercial rearings since 2009 and 2010. The AdDNV genome was cloned and sequenced for this study. The transcription map showed that splicing occurred in both the nonstructural (NS) and capsid protein (VP) multicistronic RNAs. The splicing pattern of NS mRNA predicted 3 nonstructural proteins (NS1 [576 codons], NS2 [286 codons], and NS3 [213 codons]). The VP gene cassette contained two VP open reading frames (ORFs), of 597 (ORF-A) and 268 (ORF-B) codons. The VP2 sequence was shown by N-terminal Edman degradation and mass spectrometry to correspond with ORF-A. Mass spectrometry, sequencing, and Western blotting of baculovirus-expressed VPs versus native structural proteins demonstrated that the VP1 structural protein was generated by joining ORF-A and -B via splicing (splice II), eliminating the N terminus of VP2. This splice resulted in a nested set of VP1 (816 codons), VP3 (467 codons), and VP4 (429 codons) structural proteins. In contrast, the two splices within ORF-B (Ia and Ib) removed the donor site of intron II and resulted in VP2, VP3, and VP4 expression. ORF-B may also code for several nonstructural proteins, of 268, 233, and 158 codons. The small ORF-B contains the coding sequence for a phospholipase A2 motif found in VP1, which was shown previously to be critical for cellular uptake of the virus. These splicing features are unique among parvoviruses and define a new genus of ambisense densoviruses.

Résumé

Le virus *Acheta domesticus densovirus* (AdDNV), isolé à partir de criquets est endémique en Europe depuis au moins 35 ans. De graves épizooties ont aussi été observées dans des élevages américains depuis 2009-2010. Le génome du virus AdDNV a été cloné et séquencé dans cette étude. L'étude de la stratégie de transcription a montré un épissage dans les ARNs multicistroniques des protéines structurales (VP) et non structurales (NS). Le patron d'épissage de l'ARNm NS a prédit trois protéines non structurales (NS1 [576 codons], NS2 [286 codons], et NS3 [213 codons]). La cassette VP contient deux cadres de lecture ouverts (ORF) de 597 (ORF-A) et 268 (ORF-B) codons. La séquence de VP2 a été déterminée par dégradation d'Edman et spectrométrie de masse et correspond à l'ORF-A. Les techniques de spectrométrie de masse, de séquençage et d'immunobuvardage de type western à partir de protéines VP exprimées dans un système baculovirus et de protéine structurales natives ont montré que VP1 est générée suite à la jonction des ORF A et B par épissage (épissage II), éliminant la région N-terminale de VP2. Le résultat de cet épissage permet la production de VP1 (816 codons), VP3 (467 codons), et VP4 (429 codons). Au contraire, les deux épissages de l'ORF-B (la et lb) ont retiré le site donneur de l'intron II, résultant en la production de VP2, VP3 et VP4. L'ORF-B peut aussi produire des protéines non structurales de 268, 233 et 158 codons. Le petit ORF-B contient la séquence codante du motif enzymatique de phospholipase A2 retrouvé dans la protéine VP1, qui a précédemment été démontré comme étant critique pour l'entrée cellulaire du virus. Ces caractéristiques d'épissage sont uniques chez les parvovirus et définissent un nouveau genre de densovirus ambisens.

Introduction

Insect parvoviruses (densovirus) belong to the *Densovirinae* subfamily of the *Parvoviridae* and are small, isometric, nonenveloped viruses (diameter, 25 nm) that contain a linear single-stranded DNA of 4 to 6 kb (Bergoin *et al.*, 2008, Bergoin *et al.*, 2010, Tijssen *et al.*, 2006a). These viruses can be subdivided into two large groups, those with ambisense genomes and those with monosense genomes. Like vertebrate parvoviruses, all densovirus have a genomic DNA with hairpins at both ends, often (but not necessarily for all genera) as inverted terminal repeats (ITRs). All densovirus with ambisense genomes package both complementary strands in equimolecular ratios as single strands in separate capsids (Tijssen *et al.*, 2006a). The nonstructural (NS) gene cassette is found in the 5' half of one genome strand, and the structural protein (VP) gene cassette is found in the 5' half of the complementary strand. By convention, the genome is oriented so that the NS cassette is found in the left half. Expression strategies of densovirus often involve (alternative) splicing and leaky scanning translation mechanisms (Tijssen *et al.*, 2003). So far, the near-atomic structures of three densovirus, *Penaeus stylirostris* densovirus (PstDNV), *Bombyx mori* densovirus 1 (BmDNV-1), and *Galleria mellonella* densovirus (GmDNV), have been solved (Kaufmann *et al.*, 2010, Kaufmann *et al.*, 2011, Simpson *et al.*, 1998). The capsid of densovirus consists of 60 subunits ($T=1$) of identical proteins that may contain N-terminal extensions not involved in capsid formation but that confer additional functions to the capsid. One of these functions is a phospholipase A2 (PLA2) activity that is required for genome delivery during infection (Zádori *et al.*, 2001). Densovirus are usually highly pathogenic for their natural hosts (Fédière, 2000).

The monosense densovirus have been classified into 3 uniform genera*, i.e., *Iteravirus**, with a 5.0-kb genome, 0.25-kb ITRs, and a PLA2 motif in VP; *Brevidensovirus*, with a 4.0-kb genome, no ITRs but terminal hairpins, and no PLA2 motif; and *Hepanvirus**, with a single member, hepatopancreatic parvovirus, with a 6.3-kb genome also lacking a PLA2 motif and ITRs but with 0.2-kb terminal hairpins

(Sukhumsirichart *et al.*, 2006, Tijssen *et al.*, 2006a). In contrast, the ambisense densovirus have been classified into one uniform genus, *Densovirus*^{*}, with a 6-kb genome and 0.55-kb ITRs, and a second genus, *Pefudensovirus*, with only *Periplaneta fuliginosa* densovirus (PfDNV) as a member, with a 5.5-kb genome, 0.2-kb ITRs, and a split VP gene cassette (Bergoin *et al.*, 2010, Tijssen *et al.*, 2011). Ribosome frameshifts have been proposed to connect its VP open reading frames (ORFs) (Yamagishi *et al.*, 1999). So far, all ambisense densovirus have an N-terminal PLA2 motif in their largest VP. Some sequenced ambisense densovirus, e.g., *Myzus persicae* densovirus (MpDNV) (van Munster *et al.*, 2003b), *Blattella germanica* densovirus (BgDNV) (Mukha *et al.*, 2003), and *Planococcus citri* densovirus (PcDNV) (Thao *et al.*, 2001), are yet unclassified. The ambisense virus *Culex pipiens* densovirus (CpDNV) has a different genome organization for both the NS and VP proteins and will have to be classified in a different genus (Baquerizo-Audiot *et al.*, 2009).

Acheta domesticus densovirus (AdDNV) was isolated from diseased *Acheta domesticus* L. house crickets from a Swiss commercial mass rearing facility (Meynardier *et al.*, 1977). The virus spread rapidly and was responsible for high mortality rates, such that the rearing could not be saved. This was the first observation of a Densovirus in an orthopteran species. Infected tissues included adipose tissue, the midgut, the hypodermis, and particularly the Malpighi tubules, but the most obvious pathological change was the completely empty digestive cecae (Szelei *et al.*, 2011). The cecae, which flank the proventriculus, are the sites where most enzymes are released and most absorption of nutrients occurs. Feulgen-positive masses were observed in the nuclei of infected cells (Meynardier *et al.*, 1977). Commercial production facilities for the pet industry or for research mass rearings of house crickets in Europe are frequently affected by this pathogen. This virus was previously not known to circulate in North America, except for a small epidemic in the Southern United States in the 1980s (Styer *et al.*, 1991). Beginning in 2009, sudden, severe outbreaks were observed in commercial facilities in Canada and the United States, leading to losses of hundreds of millions of dollars and to an acute crisis in the pet food

industry (Szelei *et al.*, 2011). In this study on AdDNV, we observed that over the last 34 years, the annual replacement rate was about 2.45×10^{-4} substitution/nucleotide (nt) and that the VP gene cassette consists of two ORFs, a characteristic of the *Pefudensovirus* genus (Szelei *et al.*, 2011).

In the present study, the complete genome and the expression strategy of AdDNV were investigated and showed features not yet described for other densoviruses or vertebrate parvoviruses. The most striking observation was the intricate splicing pattern of its VP ORFs, resulting, in contrast to the case for all parvoviruses studied so far, in unrelated N-terminal extensions of its two largest structural proteins and in the probable production of several supplementary NS proteins from the VP cassette.

Materials and methods

Rearing of crickets.

A. domesticus L. house crickets were obtained from a commercial supplier and were reared under conditions of about 60% relative humidity, 28°C, and a 16h–8h light-dark cycle. Diet conditions and drinking water supply, as well as conditions for perching, hiding, and oviposition, were as described previously (Szelei *et al.*, 2011). Infection techniques. The visceral cavity of nymphs of about 1.5 to 2 cm was injected with an inoculum consisting of a viral suspension obtained by grinding an infected cricket in 1X phosphate-buffered saline (PBS) plus 2% ascorbic acid, clarifying the mixture by centrifugation for 10 min at 8,000 g, and filtering it through 450-nm membranes. Mortality was usually 100% within 2 weeks. Alternatively, infection was achieved by feeding with a virus-contaminated diet as previously described (Szelei *et al.*, 2011).

Virus and DNA preparation.

Virus was purified as previously described (Tijssen *et al.*, 1977). Lysis buffer [300 µl of 6 M guanidine-HCl, 10 mM urea, 10 mM Tris-HCl, 20% Triton X-100, pH 4.4, containing 80 µg/ml poly(A) carrier RNA] and 200 µl sample were mixed and incubated for 10 min at 70°C. The sample was vortexed after adding 125 µl isopropanol, and the DNA was then purified on HighPure plasmid spin columns (Roche Molecular Biochemicals) according to the supplier's instructions.

Cloning, mutation analysis, and sequencing of viral DNA.

The 1977 isolate of AdDNV was cloned into the pCR-XL-TOPO vector (Invitrogen Life Sciences), using supercompetent *Sure 2 Escherichia coli* cells (Stratagene) at 30°C. Point mutations in the AdDNV genome were generated with a QuikChange site-directed mutagenesis kit (Stratagene), whereas deletion mutants were obtained via the type IIb restriction endonuclease strategy (Fernandes *et al.*, 2009). Independent clones were sequenced in both directions by primer walking. The terminal hairpins yielded compressions that were difficult to sequence; however, inclusion of 1 M betaine (Sigma) and 3% dimethyl sulfoxide (DMSO) or restriction in the hairpin by *Dra*I yielded clean sequence reads. DNAs from subsequent isolates were amplified by PCR and sequenced between the ITRs.

Isolation and characterization of viral RNA.

Total RNAs were isolated from 30 mg adipose tissue from infected cricket larvae (2 to 5 days postinfection [p.i.]) and from recombinant baculovirus-infected cells at 48 h p.i. by use of an RNeasy minikit from Qiagen. The DNase I treatment was extended from 15 to 30 min or repeated twice. A PCR test was included to verify the absence of DNA. Total extracted RNA was subjected to mRNA purification using an mRNA isolation kit (Roche).

Northern blots.

About 20 to 30 µg total RNA in a 6-µl volume was added to 18 µl buffer (1X MOPS [20 mM morpholinepropanesulfonic acid, 5 mM sodium acetate, 0.5 mM EDTA adjusted to pH 8 with NaOH], 18.5% formaldehyde, 50% formamide), 5 µl loading buffer was added, and the mixture was incubated for 5 to 10 min at 65 to 70°C and separated by electrophoresis on a 1% formaldehyde-agarose gel. Parallel lanes contained RNA size markers (Promega). After migration and washing, RNAs were transferred to positively charged nylon membranes (Roche) by capillary blotting overnight. The blotted membranes were prehybridized with 10 mg/ml herring sperm DNA in 50% formamide before hybridization with 32 P-labeled probes. The probes corresponded to a 1.5-kb *Bgl*II-*Sall* restriction fragment specific for the VP coding sequence and a 0.87-kb *Eco*47III-*Dra*I restriction fragment specific for NS. Hybridized probes were visualized with a Storm 840 phosphorimager.

Mapping of 5' ends, 3' ends, and introns of viral transcripts.

The most probable locations of the transcripts were predicted from the ORFs obtained by sequence analysis. A 3' rapid amplification of cDNA ends (3'-RACE) system was used to characterize the 3' ends of the polyadenylated transcripts, using the RNAtag and ADAP primers (Table 2.2) and PCR (Tijssen *et al.*, 2003), whereas the 5' ends were determined with a FirstChoice RLM RACE kit (Ambion) according to the instructions of the supplier. The locations of introns were determined after reverse transcription of the transcripts by use of avian myeloblastosis virus (AMV) reverse transcriptase (Promega) in a final volume of 20 µl for 1 h at 42°C, PCR using internal DNV-specific primers for overlapping regions, and sequencing of the amplicons according to standard methods (Tijssen *et al.*, 2003).

Promoter activity in AdDNV genome.

Promoter regions were amplified by PCR and cloned upstream of the luciferase gene in the pGL3-basic system. The ProNSF and ProNSR primers were used for the

NS promoter, the ProVP1F and ProVP1R primers were used for the VP1 promoter, and the PrNSMf and PrNSMr primers were used for the *Mythimna loreyi* densovirus (MIDNV) NS promoter (control) (Table 2.2). Sequencing was performed to confirm the promoter direction. For the assay, Ld652 cells were seeded into wells of 24-well cell culture plates. Each well contained about 0.5 ml of cells at 5×10^5 cells/ml. The cells were cultured overnight. Transfection was performed with 2.5 μ l DOTAP reagent and 0.6 g DNA in 15 μ l HEPES, and the mixture was added to 245 μ l medium (without antibiotic or fetal bovine serum [FBS]) per well. Cells were harvested at 48 or 60 h post transfection, washed twice with PBS, and resuspended in 100 μ l of Bright-Glo lysis buffer (Promega). Cell lysates were quickly centrifuged to remove cell debris, and 25- μ l aliquots of the cell extract were used to determine luciferase activity according to the instructions for a luciferase assay system (Promega).

Expression of structural proteins and analysis of VP ORFs by use of a baculovirus system.

The potential VP coding sequences were cloned into the *Autographica californica* nuclear polyhedrosis virus (AcNPV) downstream of the polyhedrin promoter by use of the Bac-To-Bac baculovirus expression system (Luckow *et al.*, 1993) (Invitrogen) via the pFastBac1 and pFastBacHT vectors according to the supplier's instructions. In constructs involving expression of VP1, the initiation codon had to be moved closer to the start of the transcript. For this purpose, an *Eco*RI site was introduced 100 bp upstream of the multiple cloning site (MCS), using the pFECRF and pFECRIR mutation primers (Table 2.2), followed by removal of the small *Eco*RI fragment between the new and MCS *Eco*RI sites. Inserts were generated by PCR (Tijssen *et al.*, 2003) with the primers given in Table 3-3, using the wild-type (wt) template or a template in which intron II (see below) splicing which coincided with the initiation codon of VP1, or an equivalent in which the initiation codon ATG was mutated to ACC (AdmATG1B), and the reverse primer Ad1HAR, containing an *Xba*I site, was used (Table 2.2). All pFastBac recombinant constructs were verified by sequencing. Protein analysis by SDS-PAGE, Western blotting, and N-terminal amino acid

sequencing. Capsid proteins were analyzed by SDS-PAGE (Laemmli, 1970), using the structural proteins of Junonia coenia densovirus (JcDNV) or broad-range standards (Bio-Rad) as size markers. Expressed proteins were analyzed by Western blotting (Tijssen *et al.*, 2003, Towbin *et al.*, 1979), using polyvinylidene difluoride (PVDF) membranes and Roche blocking reagent. For amino acid sequencing, structural proteins from AdDNV were separated by SDS-PAGE on 10% polyacrylamide gels and were electro-blotted onto nitrocellulose membranes (Westran, Schleicher & Schuell, Keene, NH) and sequenced according to the method of Matsudaira (Matsudaira, 1987).

Mass spectrometry (MS).

Expressed proteins from baculovirus constructs and native proteins from the virus were analyzed by mass spectrometry (MS) after separation by SDS-PAGE. The proteins, dissociated with 2% SDS at 95°C for 5 min, were run in a 10% acrylamide gel (Laemmli, 1970). The protein bands were cut from the gel and destained with water-sodium bicarbonate buffer and acetonitrile. Each protein was reduced with dithiothreitol (DTT) and alkylated with iodoacetamide prior to ingel digestion with trypsin (Hellman *et al.*, 1995). The tryptic peptides were eluted from the gel with acetonitrile containing 0.1% trifluoroacetic acid. The tryptic peptides were then separated on an Agilent Nanopump instrument using a C 18 Zorbax trap and an SB-C 18 Zorbax 300 reversed-phase column (150 mm 75 µm; 3.5 µm particle size) (Agilent Technologies, Inc.). All mass spectra were recorded on a hybrid linear ion trap–triple-quadrupole mass spectrometer (Q-Trap; Applied Biosys-tems/MDS Sciex Instruments, CA) equipped with a nano-electrospray ionization source. The analysis of MS-MS data was performed with Analyst software, version 1.4 (Applied Biosystems/MDS Sciex Instruments, CA). MASCOT (Matrix Science, London, United Kingdom) was used to create peak lists from MS and MS/MS raw data.

Nucleotide sequence accession number. The AdDNV sequence is available in the GenBank database under accession number HQ827781.

Results

AdDNV infection of *A. domesticus*.

AdDNV is a frequent cause of epizootics in commercial or research mass rearing facilities for house crickets in Europe. The highest mortality is observed in the last larval stage and in young adults. These crickets die slowly over a period of several days; although they appear healthy, they lie on their back and do not move. The guts of infected *A. domesticus* crickets that are still alive and no longer move are usually completely empty. Beginning in September 2009, mass epizootics have also occurred in rearing facilities throughout North America.

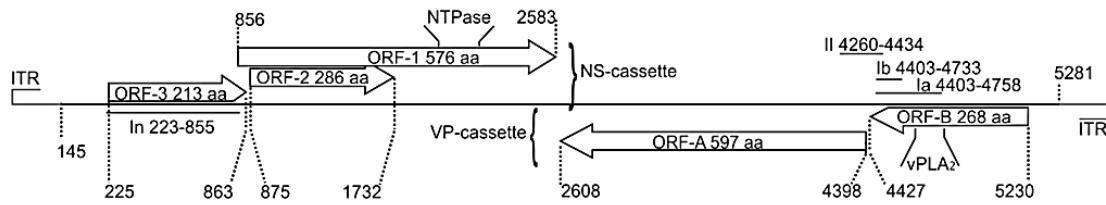
DNA sequence and organization of AdDNV isolates.

Three full-length genomic clones in the pCR-XL-TOPO vector, namely, pAd22, pAd25, and pAd35, were obtained from the 1977 AdDNV isolate. Both strands of the viral genomes were sequenced (for the full annotated sequence, see Figure S1 in the supplemental material). Nucleotide substitutions in more recent isolates have been reported elsewhere (Szelei *et al.*, 2011). The total length of the genome was 5,425 nt and contained ITRs of 144 nt, of which the distal 114 nt could fold into a perfect I-type palindromic hairpin (Figure 5.1 A, B, and D). The side arms in the typical Y-shaped terminal palindromes of many parvoviruses were missing in the case of AdDNV.

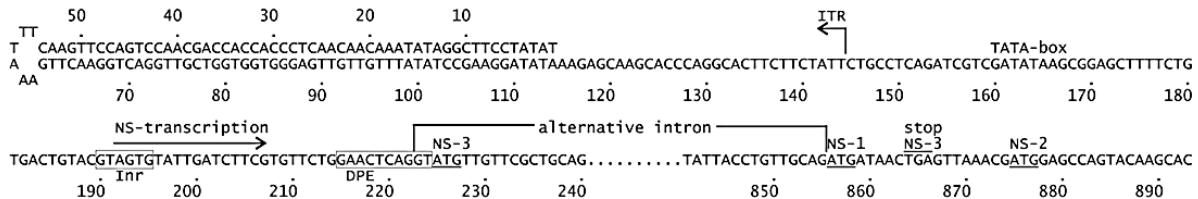
Both complementary strands contained large ORFs in their 5' halves; one strand had 3 large ORFs (ORFs 1 to 3), 2 of which were overlapping, and its complementary strand had 2 large ORFs (ORF-A and -B) (Figure 5.1A). ORFs 1 to 3 potentially code for proteins consisting of 576, 286, and 213 amino acids (aa), respectively, whereas ORFs A and B potentially code for proteins of 597 and 268 aa, respectively. nBLAST analysis (<http://www.ncbi.nih.gov>) of the 5 ORFs revealed that the ORF-1 product is a homologue of densovirus NS1 proteins, the ORF-2 product is a homologue of densovirus NS2 proteins, the ORF-A product is a homologue of densovirus VP

proteins, the ORF-B product is a homologue of parvoviral phospholipase A2 proteins (N-terminal sequence of VP1), and the ORF-3 product does not have homologous proteins. NS1 ORF-1 can also be recognized by the presence of rolling circle replication and Walker A and B motifs, and the VP ORF can be recognized by the presence of a PLA2 motif (Figure 5.1 A; see Figure S1 in the supplemental material). Since the convention for all parvoviruses is to have the genes coding for the non-structural proteins in the left half of the genome, it was decided to define the strand of the ambisense AdDNV genome containing the ORFs for the NS genes as the plus strand so the genes would be located similarly.

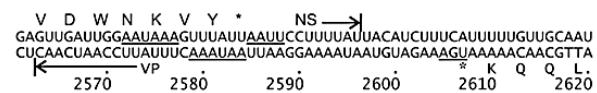
A. Genome organization: ITRs, ORFs and introns



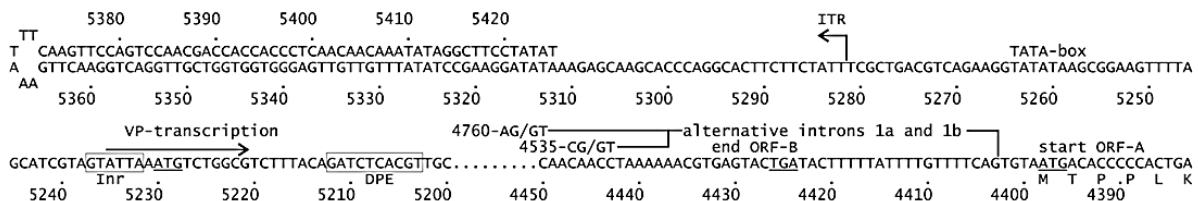
B. Left ITR, NS promoter elements and NS transcription



C. Transcription/translation ends and overlap of mRNAs



D. Right ITR, VP promoter elements, VP transcription and ORFs-A and B



E. Connection of ORF-A and ORF-B to express VP1

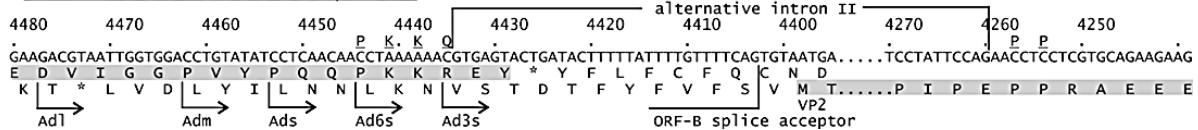


Figure 5.1 Genome organization of AdDNV.

(A) Overview of genome organization, positions, and sizes of ITRs, ORFs, and introns. The Intron In in the NS mRNA, between nt 223 and nt 855, occurs in about half of the NS transcripts. Ia (nt 4403 to 4758), Ib (nt 4403 to 4533), and II (nt 4260 to 4434) are introns that occur in alternative VP transcripts. vPLA2 indicates the position of the viral phospholipase A2 motif. (B) Left ITR and regulation of production of NS transcripts. NS transcripts start at 192-A and yield NS3 from 225-A. However, a fraction of these transcripts are spliced just upstream of this start codon (intron In), leading to translation of NS1 from 856-A (AUG with a poor initiation environment) and, through leaky scanning, of NS2 from 875-A. Inr and DPE are promoter elements. (C) Like the case for all members of the *Densovirus* genus, the 3 ends of AdDNV NS and VP transcripts overlap in the middle of the genome. The stop codons and AATAAA motifs are underlined. (D) Right ITR, VP transcription sites, and splicing in ORF-B on the complementary strand. Transcription starts at nt 5235, and VP1 initiation is at nt 5230. The short 5'-UTR predicts an inefficient initiation (leaky scanning) and could be responsible for the production of a nested set of N-terminally extended viral proteins. However, removal of either of the two alternative introns in ORF-B (Ia or Ib) did not connect the exons in ORF-B and ORF-A in frame, so only nonstructural proteins could be produced from nt 5230 and VP2 could be produced directly from the first AUG in ORF-A when this splicing occurred. (E) An alternative intron II, which is mutually exclusive with introns Ia and Ib because the ORF-B splice acceptor is removed, connects ORF-B and ORF-A (both shaded) in frame so that VP1 can be produced from nt 5230. The VP1 sequence around the splicing site is underlined and shown above the nt sequence.

SDS-PAGE revealed that the capsid is composed of 4 structural proteins with estimated molecular masses ranging from 43 to 110 kDa (Figure 2.9A), although a fifth protein may arise during purification, probably due to proteolysis (not shown). Attempts to obtain N-terminal sequences failed for VP1, VP3, and VP4, but the sequence TPPLKPHP(I)(E) was obtained for VP2, which indicated that its translation started at the AUG start codon of ORF-A, at nt 4398 to 4396 (Figure 5.1D; see Figure S1 in the supplemental material), and predicted a molecular mass of 65.3 kDa for VP2. ORF-B encoded the PLA2 motif recently identified in the structural proteins of most parvoviruses (Farr *et al.*, 2005, Fédière *et al.*, 2002, Tijssen *et al.*, 2003, Zádori *et al.*, 2001) but was too small to code for a VP1 of 110 kDa as estimated by SDS-PAGE (Figure 5.2A). Therefore, splicing of ORF-A and ORF-B seemed necessary to code for the largest AdDNV structural protein. Since the N-terminal coding region of ORF-A before its first ATG overlapped with the C-terminal coding region of ORF-B (Figure 5.1E; see Figure S1), an unspliced transcript could also code for VP1 by a ribosome frameshift in the ORF-A–ORF-B overlapping region, as suggested for PfDNV (Yamagishi *et al.*, 1999). These hypotheses were investigated further by transcript mapping and expression studies.

Northern blotting and mapping of viral transcripts.

Northern blotting of RNAs obtained from infected Acheta larvae revealed two bands of NS transcripts (about 2.5 and 1.8 kb) and one band of VP transcripts (about 2.5 kb) (Figure 5.2B). The transcript maps for RNAs isolated from both diseased crickets and recombinant baculovirus-infected cells were established by 5'- and 3'-RACE and are shown in Figure 5.1B to D. The 3' termini of the NS and VP transcripts had a 34-nt overlap (Figure 5.1C), similar to the situation observed with members of the *Densovirus* genus (Tijssen *et al.*, 2006a). NS transcription and splicing followed the same strategy as that previously described for GmDNV (Tijssen *et al.*, 2003). A large unspliced transcript (nt 192 to 2596) was found to code for NS3 (first AUG in ORF-3), starting at nt 225. The NS3 coding sequence was removed in the spliced form in roughly half of these transcripts, with an intron from 221-AG/GT to 853-CAG,

resulting in a 1,770-nt transcript that was able to code for NS1, starting at the new first codon (856-AUG), and NS2, starting at 875-AUG, by a leaky scanning mechanism due to the poor environment of the NS1 AUG codon (cagAUGa) and the strong environment of the downstream NS2 initiation codon (AcgAUGG). These two maps confirmed the sizes of the mRNAs observed by Northern blotting and indicated that this virus expressed NS1-3 in a fashion identical to that for other Densovirus members.

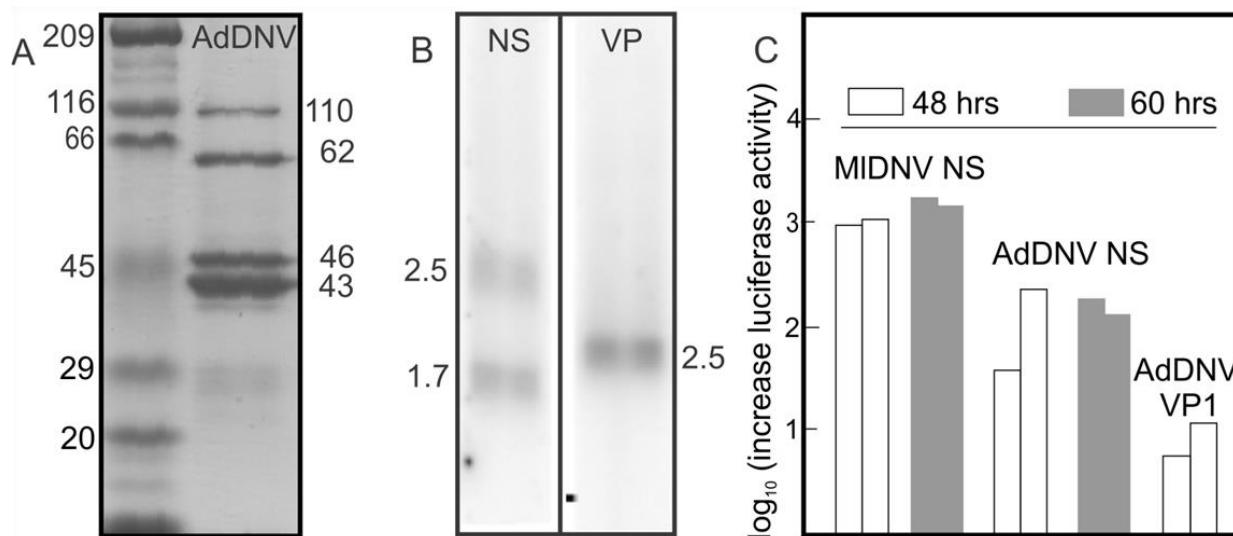


Figure 5.2 SDS-PAGE, Northern-blotting and promoter activities.

(A) SDS-PAGE analysis of structural proteins of AdDNV. Lane 1, Bio-Rad broad-range standard proteins as markers; lane 2, four proteins of AdDNV (VP1 to -4, in decreasing size). The estimated masses corresponded reasonably with the sequence-predicted masses (88, 65, 51, and 47 kDa, respectively). (B) Northern blotting of NS and VP transcripts. The two NS transcripts corresponded to a spliced and an unspliced form (see the text for further details). The single VP band consisted of at least three forms of almost identical size. (C) Promoter activities of the predicted NS and VP promoters in two independent experiments using the pGL3 vector, lepidopteran LD cells, and the Promega luciferase assay. The NS promoter was assayed at both 48 and 60 h. The NS promoter of MIDNV, a lepidopteran virus, was used as a control in the lepidopteran cells.

Table 5.1 PCR primers used in this study

Primer	Sequence ^a	Position (nt) in AdDNV	Target or use
AdVPR	TTTGTTGCAATCCCATAATAGTAC	2610–2633	Near 3 end of VP mRNA
NAdR	gctctagatCATCTGAACGTTACCACACT	3892–3915	Just upstream of VP4
Adsp	tcggaattcCACGTTCTGTGGATGAGG	4362–4380	19–37 nt into VP2
Adsvp	gccTACCAAGAAATCCGTGTAATGACA	4546–4534/4403–4393	Small intron splice
Ad3s	CGTGAGTACTGATACTTTTATT	4435–4412	End of ORF-B (TGA)
Ads	gccCCTCAACAAACCTAAAAACGTGAGTACTGA	4453–4424	End of ORF-B (TGA)
Ad6s	CCTAAAAAACGTGAGTACTGA	4444–4424	End of ORF-B (TGA)
Adl	gccGACGTAATTGGTGGACCTGTATATCCT	4477–4451	End of ORF-B (8 aa)
Adm	gccCCTGTATATCCTCAACAAACCTAAA	4462–4439	End of ORF-B (4 aa)
AdlgF	tcggaattcATGTCTGGCGTCTTACA	5230–5213	Start of VP1 (ORF-B)
RNAtag	gggtcttagagtcgagT ₁₇	Poly(A) sequence	3'/5'-RACE (first round)
ADAP	gggtcttagagtcgagT		Subsequent rounds of RACE
ProNSF	acgttaccGATATAAAGAGCAAGCACCC	109–128	NS1 promoter (<i>Kpn</i> I)
ProNSR	gaagatCTGTGCTGGAGGCCTCTACTGCAGCGAACAA CGTACCTGAGTCCAGAACAC	240–207	NS1 promoter (reverse)
ProVP1R	gaagatctAAGACGCCAGAGATTTAACT	5217–5238	VP1 (<i>Bgl</i> II site; reverse)
ProVP1F	gtaggttaccGATATAAAGAGCAAGCACCC	5323–5300	VP1(<i>Kpn</i> I site; forward)
PrNSMf	ACGGTACCGACTATAATAGAGCTGAGC		MIDNV (forward)
PrNSMr	GAAGATCTATCTTGCATAGATATACCTA		MIDNV (reverse)
pFECRIF	CGCAAATAATAAGAATTCTACTGTTTCGTAAC		Mutation in pFastBac1
pFECRIR	GTTACGAAAACAGTAGAATTCTTATTATTGCG		Mutation in pFastBac1
Ad1HAR	GCTCTAGATCAAGCGTAATCTGGAACATCGTATGGGTA TTTTGTTGCAATCCCATAATA		VP products (reverse primer HA)
AdATG1B	ggaattcATGCCGTCTTACAGATCTCAC ggaattcACCTCCGTCTTACAGATCTCAC		
AdmATG1B			

^aThe AdDNV sequence is shown in capital letters. Underlined nucleotides indicate stop codons (e.g., TGA) or restriction sites (e.g., ggttacc [*Kpn*I]).

The single band of VP transcripts observed by Northern blotting could actually represent different forms of mRNA with similar intron sizes. The unspliced form, starting at nt 5235, would be 2,672 nt [plus the poly(A) tail] long. First, we determined whether ORF-A and ORF-B could be connected in frame by splicing, using RT-PCR with primers Adsp and AdlgF (Table 5.1). For all parvoviruses studied in this respect, VP1 is an N-terminally extended form of VP2, and the position of the Adsp primer was therefore chosen about 25 nt downstream of the VP2 start codon in ORF-A, with AdlgF located at the start of ORF-B. Two alternative introns, with sizes of 131 and 356 nt, were found with the same splice acceptor at codon 4405-CAG (Figure 5.2D and 5.3A and B). Both introns failed to yield an in-frame coding sequence with ORF-A. The stop codon in the spliced ORF-B overlapped the start codon for VP2 (ugUAAUGa) (Figure 5.1D). In some systems, e.g., influenza B virus (Powell *et al.*, 2008a, Powell *et al.*, 2008b) and some non-long-terminal-repeat (non-LTR) retrotransposons (Kojima *et al.*, 2005), reinitiation occurs at such stop-start sequences. Expression of the intronless sequence from nt 4546 (before the small intron) to nt 3892 in the baculovirus system via pFastbacHTb, using primers Adsyp and NAdR (Table 5.1) and cloning using *Ehel* and *Eco*RI sites, did not yield products larger than the 29 aa expected from the baculovirus/ORF-B construct, arguing against reinitiation. These results gave credence to the previously suggested ribosome frameshift for PfDNV to generate a nested set of N-terminally extended structural proteins, as observed for all parvoviruses studied thus far (Yamagishi *et al.*, 1999). To test this hypothesis, we made several recombinant baculovirus constructs such that their expression products would be amenable to N-terminal sequencing in the potential frameshift region and they would be of reasonable size for mass spectrometry (Figure 5.1E).

Expression of VP1.

Figure 5.1E shows the 43-nt overlap of ORF-B with the N-terminal extension of ORF-A and the positions of primers (after the ORF-B and before the ORF-A stop codon) used to study the potential translational frameshift. The pFastbacHTb vector was used to yield products that could be purified via their N-terminal His tails and cleaved with

the tobacco etch virus (TEV) protease, leaving only one codon (Gly) upstream of the insertion at the blunt-end *Ehel* restriction site (Figure 5.4A). The PCR products obtained using the forward primers Ad3s, Ad6s, Ads, Adm, and Adl (Table 5.1) (the distance from the ORF-B stop codon is indicated, in codons, in Figure 5.4A) and the reverse primer NAdR, chosen at the beginning of the extension of VP4 to ensure stable products, were cloned into the *Ehel* site, and the clones were selected for orientation by *Xba*I analysis (the *Xba*I sites in the MCS and the NAdR primer should be close to each other) and then verified by sequencing.

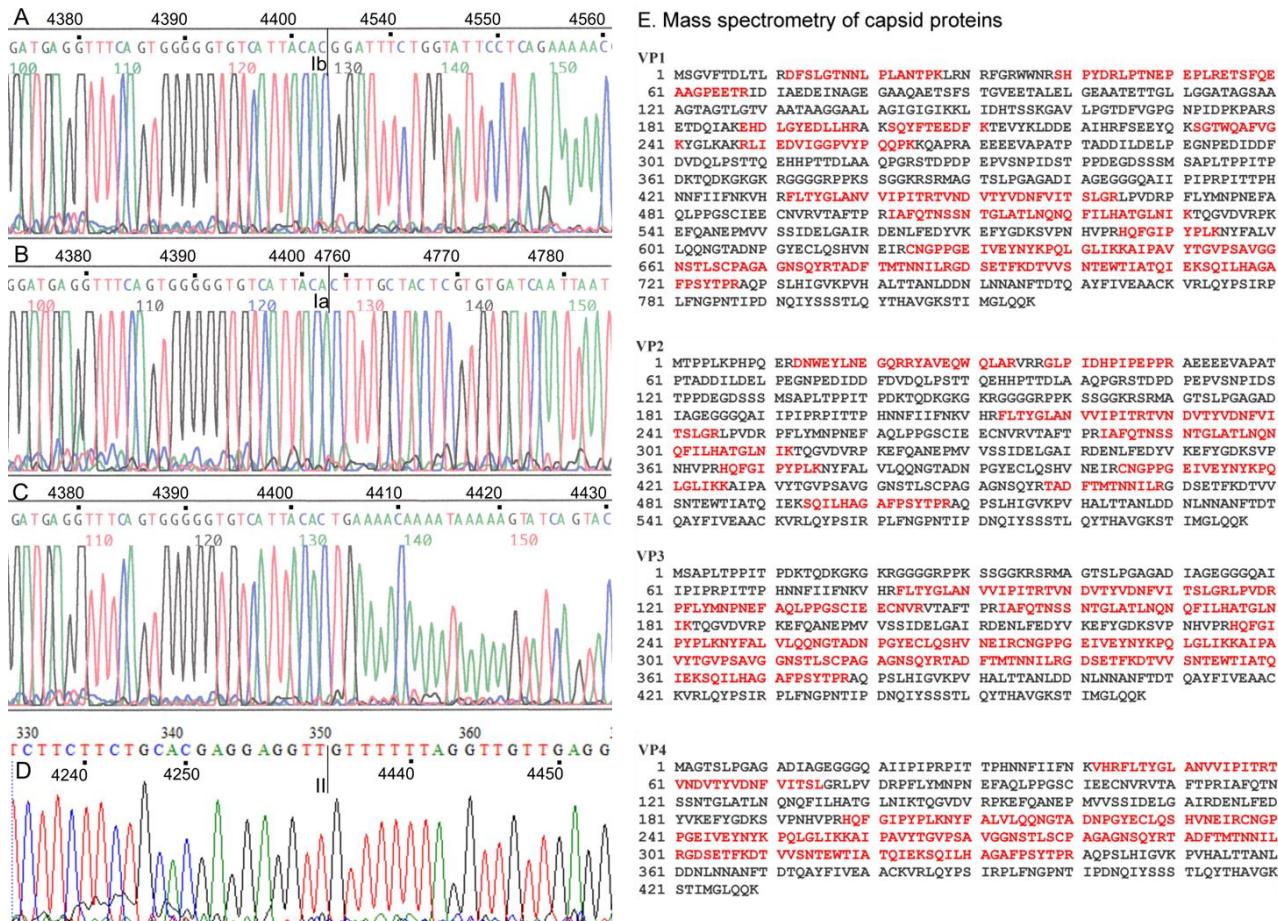


Figure 5.3 Results for mRNA sequencing and mass spectrometry.

(A to D) Sequences of mRNAs in the VP ORFs, i.e., removal of the Ib intron in ORF-B (A), removal of the la intron in ORF-B (B), unspliced mRNA (C), and the connection of ORF-A with ORF-B by splice II (D). (E) Mascot search results after mass spectrometry for capsid proteins from viruses in natural infections. The proteins were treated with trypsin, which cuts at the C-terminal side of K or R unless the next residue is P. Matched peptides are shown in bold red.

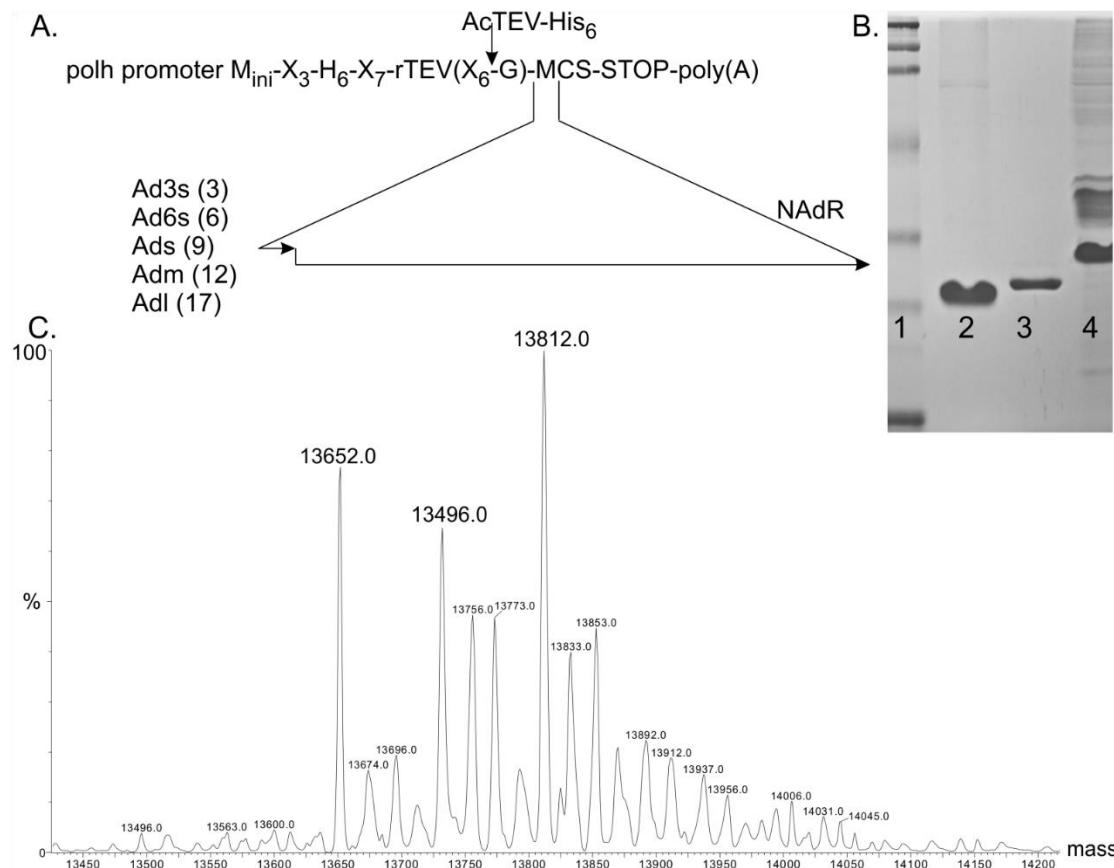


Figure 5.4 Conformation of the splicing between ORF-A and ORF-B.

(A) Construction of various ORF-A and ORF-B constructs with different Ad primers (Ad3s, Ad6s, etc.) linking a short part of ORF-B to a larger part of ORF-A, using the NAdR primer and the pFastbacHTb baculovirus system. Upstream of the insert is a TEV proteolytic site and a histidine tail (H6); N-terminal sequencing of the digested purified protein would reveal the connecting sequences. The numbers in parentheses with the different primers (Ad3s, Ad6s, etc.) denote the distances in codons from the ORF-B stop codon (small arrow). (B) SDS-PAGE analysis of recombinant Ads3 protein generated with Ad3s and NAdR primers, within ORF-B and ORF-A, respectively, using the pFastbacHTb baculovirus system. Lane 1, protein markers; lane 2, Ads3 protein after AcTEV protease digestion and purification on a Ni column; lane 3, AcTEV protease (contains a His tag for easy removal after the reaction); lane 4, Ads3 protein prior to cleavage with AcTEV. (C) Mass spectrometry of TEV-treated and purified Ads3 protein, which was predicted from its sequence to have a mass of 13,650.0 Da. A major peak at 13,652 Da verified the predicted sequence. Two other major peaks, at 13,732 Da and 13,812 Da, had an additional 80 and 160 Da, indicating the addition of one and two phosphate groups, respectively. Minor bands had one or more protons replaced by sodium, each adding an additional 22 Da.

All constructs (His 6 -TEV recognition site-*Ehel* site-insert-*Xba* site) yielded proteins that could be purified via their His tails and that had molecular masses indicating that the two frames were connected. The Ad3s and Ads products were treated with TEV protease and were subjected to N-terminal sequencing after the N-terminal His tail/TEV site fragments and the protease were removed by affinity chromatography on Ni-nitrilotriacetic acid (Ni-NTA) columns and analyzed by SDS-PAGE (Figure 5.4 B). The sequences obtained (Ad3s sequence, GQP?RAEE; and Ads sequence, GAPQQP??QPPaAE), containing an N-terminal G and GA that were introduced via the primers and remained after TEV cleavage, indicated that nt 4435 in ORF-B was connected in frame with nt 4259 in ORF-A. Mass spectrometry analysis of the purified Ads product yielded a mass of 13,652 Da, with masses of 13,732 and 13,812 Da for phosphorylated species, for a predicted mass of 13,649.64 Da for the sequence GAPQQP...GGKRSR (Figure 5.4C). These results confirmed the occurrence of a splice between nt 4435 and nt 4259. The predicted mass of VP1 is thus 88 kDa less than that estimated by SDS-PAGE (Figure 5.2A), which may be explained by the phosphorylation observed by mass spectrometry.

Splicing was further investigated by RT-PCR of mRNAs extracted from infected crickets, using 2 sets of primers, namely, AdIgF/NAdR and Ads/AdVPR (Table 5.1), covering the whole coding sequence of VP1 except for the common VP4 sequence, with estimated products of 1,357 and 1,847 bp without splicing and 1,182 and 1,672 bp after intron II splicing, respectively. The intron II splice was also confirmed by sequencing of the VP1 cDNA (Figure 5.3D). As illustrated in Figure 5.1D and E, the Ia or Ib intron and the intron II splice were mutually exclusive, i.e., the intron II splice removed the acceptor site for the two intron I splices, whereas the intron I splices removed the donor site for intron II splicing. This expression strategy was further confirmed using recombinant baculovirus constructs. The VP1 sequence from which intron II was removed, and which was thus rendered resistant to ORF-B splicing (Ia and Ib), did not yield VP2 (Figure 5.5A). In contrast, constructs with a mutated VP1 initiation codon and a normal VP template yielded a strong VP2 band but also some VP3 and VP4 (Figure 5.5A), because type II splicing could remove the VP2

initiation codon and thus allow downstream initiation. When the template without intron II was used in combination with the mutated VP1 sequence, it yielded, as expected, VP3 and VP4 only. The leaky scanning of the VP3 initiation codon can probably be explained by its weak environment (uccAUGa), in contrast to the strong environment of the VP4 initiation codon (agaAUGg). Therefore, the VP multicistronic cassette yielded 2 sets of structural proteins (Figure 5.5 B).

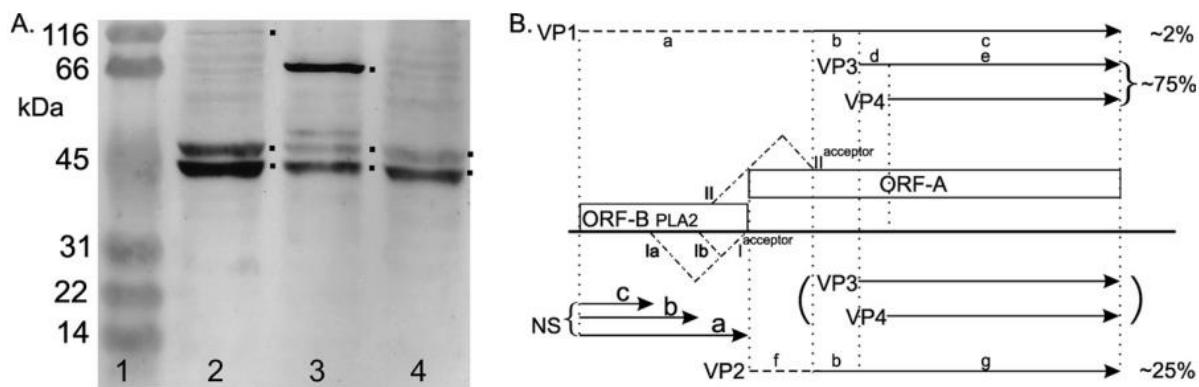


Figure 5.5 Expression of pFastbac VP constructs.

(A) Western blot of pFastbac VP constructs. Lane 1, prestained SDS-PAGE protein standards (Bio-Rad). Lane 2, expression products from the ORF-A-ORF-B construct from which intron II was deleted. Strong bands were observed for VP3 and VP4, with a weak band for VP1 (all indicated with dots), in addition to some spurious bands. Lane 3, expression products from the ORF-A-ORF-B construct with a mutated VP1 initiation codon. A strong band was observed for VP2, but relatively strong bands were also seen for VP3 and VP4, since some type II splicing occurred that removed the VP2 initiation codon. A weak band was also observed just above VP3. Lane 4, ORF-A-ORF-B construct with a mutated VP1 initiation codon and with intron II deleted, yielding only VP3 (weak initiation codon) and VP4. This lane also contained some of the spurious, nonspecific bands seen in lane 2. (B) Schematic representation of expression products of the VP cassette. Depending on the splice (Ia and Ib versus II), different sets of structural proteins (VP1, VP3, and VP4 versus VP2) were generated. VP3 and VP4 only were expressed in specific constructs (see panel A, lane 4). In addition, expression products solely from ORF-B (NS_a, NS_b, and NS_c) could be expected if initiation at the VP1 codon was combined with Ia or Ib or if no splicing occurred. The predicted pls of the common protein and N-terminal extensions differed considerably (a, 4.90; b, 3.55; c, 8.58; d, 11.60; e, 6.21; f, 6.72) and may have been a factor in the difference in observed and predicted masses of the capsid proteins (see Figure 2.9 A). The dashed N-terminal extensions of VP1 and VP2 denote their unique sequences.

Analysis of promoter activity.

The promoter elements as well as the poly(A) signals were predicted by the mapping of transcription starts and polyadenylation sites of both NS and VP transcripts (Figure 5.1B to D). To assess and compare their functionality, promoter regions (including the start of transcription) were amplified by PCR and cloned upstream of the luciferase gene in the pGL3-basic system. Their relative activities were determined by luciferase assays in independent duplicates at either 40 or 60 h posttransfection. The activity of the NS promoter of AdDNV was very significant in Ld652 cells from gypsy moths but was lower than that of the NS promoter of MIDNV, a lepidopteran densovirus (Figure 5.2C). However, the VP1 promoter activity was significantly lower, suggesting the need for trans-activation, the absence of a critical factor reacting with the non-ITR region of the VP1 promoter, or differences in transcription factors between the cricket and gypsy moth systems.

Mass spectrometry of AdDNV capsid proteins.

AdDNV was purified and the proteins separated by SDS-PAGE and subjected to mass spectrometry analysis in order to confirm the results obtained by analyzing the baculovirus constructs of the viral proteins. The proteins purified from the gel were digested with trypsin, and sequences of the peptides were determined. Analysis of VP1 and VP2 confirmed the results obtained with the baculovirus expression experiments with respect to the inframe linking of ORF-B and ORF-A. The peptide sequences obtained covered 33% of VP1, 26% of VP2, 50% of VP3, and 42% of VP4 (Figure 5.3E). One outlier peptide identified for VP3 was found in VP2, but with an ion score of 4, this was considered background.

Splicing of the Ia and Ib donor sites with the intron II acceptor site would theoretically also be possible and would yield additional products of 783 and 708 amino acids, close to the 816 amino acids observed for VP1. However, only 4 structural proteins were observed, and mass spectrometry demonstrated the presence of two

VP1-specific peptides located in the introns of these potential supplementary products, of 708 and 783 amino acids (Figure 5.6), confirming that the 816-aa species corresponded to VP1.

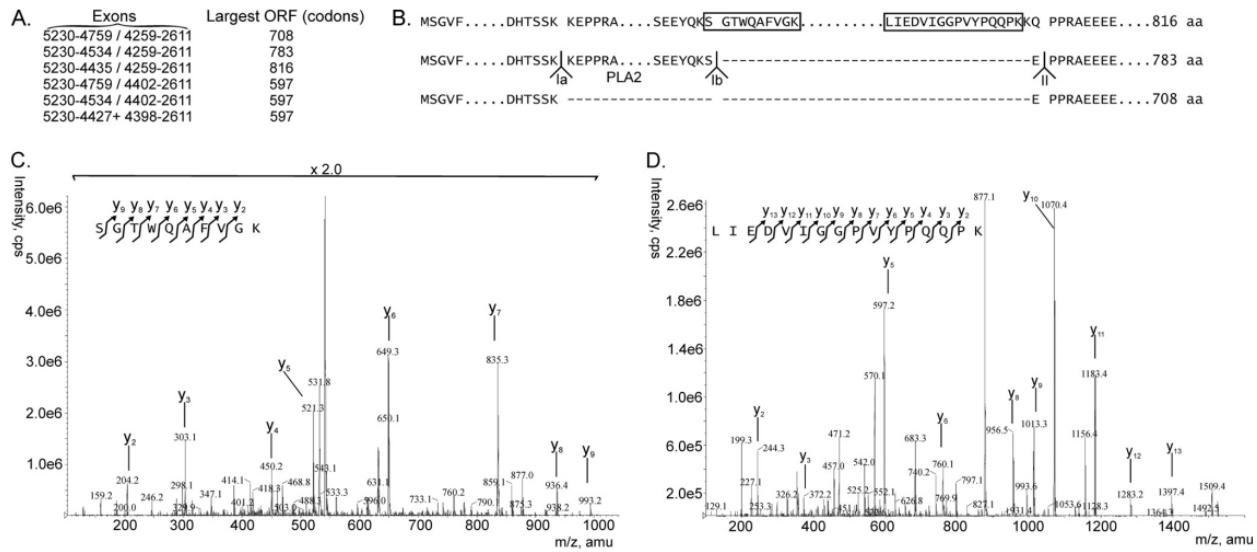


Figure 5.6 Potential VP1 proteins encoded by the VP gene cassette.

(A) Different potential splice combinations of the ORFs in the VP gene cassette and the largest protein that could then be coded. N-terminal amino acid sequencing showed that the 597-codon ORF corresponded to VP2 and, theoretically, that the three larger ORFs could code for the larger VP1 protein. (B) Schematic representation of potential sequences of large structural proteins, using the three splice combinations Ia donor-II acceptor (708 aa), Ib donor-II acceptor (783 aa), and II donor-II acceptor (816 aa). Mass spectrometry of VP1 proteins from viruses obtained after natural infections revealed the boxed sequences, which were not present in the potential 783-aa and 708-aa proteins. Therefore, the 816-aa protein corresponded to VP1. (C) Mass spectrometry identifying the unique peptides in the boxed sequences of the 816-aa protein in panel B as part of VP1. Peptides are shown with the charged C-terminal lysine, e.g., y 2 is GK (m/z 204.2), y 3 is VGK (m/z 303.1), etc., for the SG~~T~~WQAFVGK peptide. The ion score of this peptide is 61, indicating an excellent matching of the MS/MS spectrum to the identified peptide. (D) Mass spectrometry of the other peptide, similar to the peptide in panel C. The ion score is very high for this peptide, namely, 87. A detailed analysis of the peptides in panels C and D is given in Figure S2 in the supplemental material (Annexe 2).

Discussion

The 1977 isolate of AdDNV was cloned and its expression strategy analyzed. Additional AdDNV isolates from Europe, isolated in 2004, 2006, 2007, and 2009, and from North America, isolated in 1988 (Tennessee) and 2009 (Quebec, Alberta, British Columbia, and Washington State), were amplified by PCR targeting the region between the ITRs and then sequenced (reported elsewhere (Szelei *et al.*, 2011)). All 2009 North American isolates had identical sequences, suggesting a common source, and differed from the 1977, 2004, and 2006 isolates by 49, 18, and 16 nt substitutions, respectively. The genome organizations of these isolates were identical.

The sequences of the AdDNV ORFs were compared, using nBLAST, with those of other densoviruses, in particular with those of viruses such as PfDNV, PcDNV, BgDNV, MpDNV, and *Dysaphis plantaginea* densovirus (DplDNV), which also have split VP ORFs (see Figure S3 in the supplemental material). Surprisingly, the highest identities by far for the AdDNV NS1 and ORF-A proteins were found for proteins of PcDNV from *Planococcus citri* (a citrus mealybug belonging to the *Pseudococcidae* family of the Hemiptera insect order, whereas *Acheta domesticus* belongs to the *Gryllidae* family of the insect order *Orthoptera*).

The VP transcript was found to start 23 nt upstream of the 3'-ITR, at nt 5467, and the starts of both NS transcripts were at nt 573, 23 nt downstream of the 5'-ITR. This suggested that many promoter elements would be located within the ITRs and would be identical for the VP and NS promoters. The sequence context of both starts corresponded reasonably well with the consensus sequence for Inr boxes (TCAGTG); however, the promoter activity in lepidopteran cells differed considerably (Figure 5.2C). The region from the 5'-untranslated region (5'-UTR) in the VP mRNA to the putative VP1 AUG was only 5 nt long, whereas for the two NS transcripts, the 5'-UTRs were 32 (1.8-kb transcript) and 30 (2.5-kb transcript) nt long.

The expression strategy of the NS cassette is identical to that for the other members of the *Densovirus* genus. In the unspliced transcript (Figure 5.1A), NS3 is translated, whereas in the spliced form the ORF for NS3 is removed and translation starts at the weak initiation codon of NS1 or, due to leaky scanning, at the coding sequence for NS2, 19 nt downstream.

In contrast to this conserved strategy, VP expression has unique features that so far have not been observed for vertebrate parvoviruses and densoviruses, which all have a perfect nested set of N-terminally extended structural proteins. AdDNV displays split VP ORFs, and its two largest structural proteins have different extensions to which no roles have yet been assigned. PfDNV (Yamagishi *et al.*, 1999), PcDNV (Thao *et al.*, 2001), BgDNV (Kapelinskaya *et al.*, 2008, Mukha *et al.*, 2006), and MpDNV (van Munster *et al.*, 2003a, van Munster *et al.*, 2003b), which all have a split VP ORF, with the smaller ORF encoding the PLA2 motif, probably have similar expression strategies (Tijssen *et al.*, 2006a). For PfDNV, a large number of donor and acceptor splicing sites have also been identified in the VP ORFs. Splicing of the sd3 and sa3 splicing sites in cDNA11 of PfDNV could also yield a large VP1 protein. Like the case for AdDNV, many of the splicing combinations could be inconsequential for generation of the structural proteins. For BgDNV (Dmitry V. Mukha, personal communication), the larger of the two VP ORFs also codes for VP2, and splicing that is slightly different from the strategy in AdDNV also brings the two ORFs in frame and codes for VP1.

Furthermore, it was expected that initiation at the VP1 start codon, which was leaky due to the short 5'-UTR, did not depend on the presence of a downstream splice and that the various spliced and unspliced forms of mRNA were equally probable to be translated. This would lead to premature termination of translation in the case of unspliced mRNA (Figure 5.3C), yielding a 268-aa ORF-B product (NSa) that is not involved in capsid formation (Figure 5.1A). The 2 ORF-B introns would be responsible for 2 additional, minor nonstructural proteins, of 233 (NSb) and 158 (NSc)

aa. Splicing of the intron II acceptor site with the Ia and Ib donor sites was not observed (Figure 5.6).

The genome organization and expression strategy of AdDNV place it in a separate genus from *Densovirus*, and probably from *Pefudenvirus*. Unfortunately, a definitive description of the VP expression of PfDNV is still lacking to resolve whether these viruses should be coclassified. The NS cassette structure of AdDNV, PfDNV, BgDNV, and PcDNV is identical to that of members of the *Densovirus* genus but different from that of the CpDNV ambisense densovirus (Baquerizo-Audiot *et al.*, 2009), which has been proposed to be classified in a new genus, *Cupidenvirus* (Tijssen *et al.*, 2011). However, these viruses with split VP ORFs are unique among all ambisense densoviruses studied so far in that their ITRs form perfect hairpins of about 150 to 200 nt and they have genomes of about 5,450 nt (Tijssen *et al.*, 2006a). The low sequence identity among the AdDNV-like viruses may not be evocative of a need to classify them into diverse genera.

Acknowledgments

P.T. acknowledges support from the Natural Sciences and Engineering Research Council of Canada. We thank the professionals in the cricket-rearing industry for their work.

Supplemental Materials

TATATCCTCGGATATAACAAACAACCTCCCACCAAGCAACCTGACCTTGAACTTAAAGTTCAAGGTCAAGGTCAGGTTGCTGGTGGGAGTTGTTTATATCGAAGGATATAAGAGC
 10 20 30 40 50 60 70 80 90 100 110 120

ITR Start transcription NS3
 AAGCACCCAGGCACCTCTCTATTCTGCCTCAGATCGTCGATATAAGCGGAGCTTCTGTGACTGTACGTAGTGTATTGATCTCGTGTCTGGAACTCAGGTATGTTGCTGCAG
 130 140 150 160 170 180 190 200 210 220 230 240
 * alternative splicing/donor site
 10 20 30 40
 E A P P A C T F E P R S L Q Q L S L R T F V E N T D W D V I N F L Y K N N E F S
 TAGAAGGCCCTCCAGCATGACGTTGAAACCTCGTTCATTGCAACAGTTACACTTAGAACATTGAGAAAATCGGATTGGGACGTAATAAATTCTCTATAAAAATAATGAATT
 250 260 270 280 290 300 310 320 330 340 350 360
 50 60 70 80
 S R L Y S S M F Y E R K Y L D N P H E Y H D W S V P N N V P I E T Y I K I M N Y
 CTAGCAGGTTGATTCATCGATGTTTATGAAAGAAAATCTGGATAATCCGATGAATATCATGATTGGTCACTACAAATAATGTTCCATTGAAACGTAACATAAGATAATGAATT
 370 380 390 400 410 420 430 440 450 460 470 480
 90 100 110 120
 H M E D G I P S F A F Q R N L V N R C F I Q V H F C A P G T F G F P I T D N I C
 ATCACATGGAAGCGGATTCCATCATTGCTTCAACGCAATTAGTGAATAGATGTTCATACAAGTCGATTCTGCTCTGGTACTTTGGCTCCATTACAGATAATTA
 490 500 510 520 530 540 550 560 570 580 590 600
 130 140 150 160
 I P C Y N T F A R F N S S N F Y R F K K F V L V R D Q T V I E D D E L I D Y M Q
 GTATACTTGTATAACATTTGCGAGGTTAACGCTCTAAATTCTATCGTTAAAGAGTTGTTCTCGTCCGATCAAACAGTTATGAGATGATGAACTTATCGATTATGC
 610 620 630 640 650 660 670 680 690 700 710 720
 170 180 190 200
 C R S N W C T C C N T T S L F A I K D G T H L G P Q W F N S I Y L M C N E I P L
 AATGTAGAAGTAATTGGTACTTGTGAAATACAACATCATTTCGATTAAGATGAACTCATTAGGCCGAGTGGTTAATAGTATATCTCATGTAACGAAATCCGT
 730 740 750 760 770 780 790 800 810 820 830 840
 L P V A D D N * NS2 10 20 30
 210 NS1 M E P V Q A Q D N L D G V F E T L K D F M K T L W G L C E
 M I T E L N D G A S T S T G Q S G W S R G D V E G L Y E D I M G P L C R
 TATTACCTGTTGAGATGATACTGAGTAAACGATGGAGCCAGTACAAGCACAGAACATCTGGATGGAGTAGAGGAGACGTTGAAAGACTTTGAAGACATTGGGCTTCGCA
 /\acceptor site alternative splicing 850 860 870 880 890 900 910 920 930 940 950 960
 30 40 50 60 70
 N Q P L T N P Q W W A D M L E S V T I S L E D Q N T N L A S D Y A Q L K K S F L
 E S T V D Q S A M V G G H A G E C Y H F S G G S E H E P S L R L R T I K E K L P
 GAATCAACCGTTGACCAATCCGCAATGGGGCGACATGCTGGAGATGTTACCTCTGGAGGATCAGAACACGAACTTCAGCTCAGATTACGCACAAATTAAAGAAAAGCTTCC
 970 980 990 1000 1010 1020 1030 1040 1050 1060 1070 1080
 70 80 90 100 110
 K I I S H W Q T M S T K W L V T S F A D L D R R V S A L L G G T F Q T L W S R R
 E D Y I A L A D N V N E M V G H F I R G L R S K S F S S T R R Y I S D V V V T S
 GAAGATTATCGCACTGGAGACAATGTCACGAAATGGTGGTCACTTCATTGGGACTTAGATCGAGAGTTCACTCGGGTACATTCAGACGTTGAGCTCG
 1090 1100 1110 1120 1130 1140 1150 1160 1170 1180 1190 1200
 110 120 130 140 150
 L L N T V A E C W T S L I S A Q A L T P E D Y Y S G S T K E T T S T S S T T A R
 S P E H G R R V L D F L N K C A S T Y P G R L L W V N E G D H I H V V H D C P
 TCTCTGAACACGGTGCAGAGTGTGACTCCCTAAATAGTGCGCAAGCAGTACCTACCGGAAGACTACTCTGGGTCACGAAGGAGACCACATCCACGTCGACGACTGGCC
 1210 1220 1230 1240 1250 1260 1270 1280 1290 1300 1310 1320
 150 160 170 180 190
 T A L D N V D A S S Q K R K T S K R D F A S L Y A S R S S L P N S M T P T G Q M
 Y S A G N C R C K F S K T E D F K K G L R P D T E L D D T D W A N
 TACAGCGCTGACATGAGATGCAAGTCTCAAACCGGAAGACTCTAAAGGGACTTCGCAAGCCTCTAGCAAGTCGAAGTTCATTACCGAAGCTCGATGACACCGACTGGCAA
 1330 1340 1350 1360 1370 1380 1390 1400 1410 1420 1430 1440
 190 200 210 220 230
 F S Y T S L C Q N G K A N E E Q F G L M E P Y D D Y R I Q V K V Y D G P R C A S N
 V F L Y F I V S K W K S E R A V W I D G A L R R L P D T G E V G V R W T A L C E Q
 GTTTCTTACCTTGTGTCAAAATGGAAAAGCGAACGAGCAGTTGGATTGATGGAGGCTTACGAGCAGTACCGGATACAGGTGAAGGTTACGATGGACCGGGCTGTGGAGCAA
 1450 1460 1470 1480 1490 1500 1510 1520 1530 1540 1550 1560
 230 240 250 260 270
 P D Y Y W N N K L Q E M D V T V S K K S Q I W K D I N N A F C V A Y T N L E E N
 S R L L L E Q Q A A R D G C D G I E E V T D L E G H Q Q R V L R G V Y Q P R R K
 TCCCGACTACTGGAAACAACAGCTGCAAGAGATGGATGACGGTATCGAAGAAGTCACAGATTGGAGGACATCAAACCGCTTGGCTGCGTATACCAACCTCGAAGAAAA
 1570 1580 1590 1600 1610 1620 1630 1640 1650 1660 1670 1680
 270 280 290 300 310
 G A S S R P S P K K Q K Y Y S Q N *
 R S K F Q T F A E E A K V L F S K L N C I P V R D A R I I I D R K H P M Y F E L
 CGGAGCAAGTCCAGACCTTCGCCGAAGAAGCAAAGTACTATTCTCAAATTGAACTGTATTCTCTGGCTGACGCTAGAATAATAATTGATAGGAAGCATCTATGTTGAATTG
 1690 1700 1710 1720 1730 1740 1750 1760 1770 1780 1790 1800
 320 330 340 350
 H D P K N E K L Y T A A I S L Y V R D I N E M S L N D F Y L Q T L E N T C V Y Q
 CATGATCCAAAATGAAAGTTGATACTGCTGCTATTCTTATGTCGCTGACATTAATGAAATGTTCACTACAGACTTTAGAGAATACATGTGTTTATCAA
 1810 1820 1830 1840 1850 1860 1870 1880 1890 1900 1910 1920

A G N I D P F V F Y H N M E D S V H F L N E L I L F Q C N G S E E G V K E L L M
 GCTGGTAATAGACCTTTGTTTATCATATAATGGAAGATGTGTCATTCTTAATGAACTTATCTTAACTGAACTTATGTAATGGTAGGAGGAGTGTGAAAGAATTACTAATG
 1930 1940 1950 1960 1970 1980 1990 2000 2010 2020 2030 2040
 N I V A W F N R L G W Y M P T I N M N G E K E Y Q L N P K V N T V C I G N H N
 AATATTGAGCTGGTTAATAGGTTAGGTTATGCAACAAATTATGAATGGGGAAAAAGAATATCAGCTTAACCCCTAAAGTAACTGTATGATAATTGGTAATCATAATG
 2050 2060 2070 2080 2090 2100 2110 2120 2130 2140 2150 2160
 C G K N Y F W D A V C C L G N V G F L G R V N N K T N K F A L Q D C V H K R I
 TGTTGAAAGAATTATTTGGACGAGCTTGTGTTGGGTTAAATGTGGTTCTGGTAGAGCTTAAATAATAAAACTAACAGTTGCTTGCAAGATTGTTGACAAACGAA
 2170 2180 2190 2200 2210 2220 2230 2240 2250 2260 2270 2280
 V I G N E I S L E D G A K E D M K K L C E G C P F N V S V K H Q S D G I V A R T
 GTAATTGGTAATGAAATATCTTGGAAAGATGGCAAAGGAGGATATGAAAGAATTATGTGAGGGGTGTCATTCAATGTTAGCGTAAACATCAATCAGATGAA
 2290 2300 2310 2320 2330 2340 2350 2360 2370 2380 2390 2400
 P V C L I S N N C I P I A C D A H F V D V R M K V F Y W R P C S Y L A K S R K K
 CCTGTTGCTTAATTAGAACATTGTACCTATTGATGTGTCATTGGTAGATGTAAGGATGAAAGGTTATTGGAGGCCATGTAGTTGGAGGCCATGAGCTTGG
 2410 2420 2430 2440 2450 2460 2470 2480 2490 2500 2510 2520
 P Y P P A V F E L M K M Y G V D W N K V Y * end NS transcript
 CGGTATCCACCGGCTGTTTGAATTAATGAAAATGATGGAGTTGGAAATAAGTTATTACATCTTGCATTACATCTTGCATTACATCTTGCATTACATCTTGC
 2530 2540 2550 2560 2570 * 2580 2590 2600 * K Q Q L G M I T S K G
 end VP transcript 590
 2650 2660 2670 2680 2690 2700 2710 2720 2730 2740 2750 2760
 AACTGCATGAGTGACTCCAAGTAGAAGTAACTGGATAATCTGGTTATCAGGAATAGTATTAGGACCGTTAAAGAGGCTTAATGGAAAGGATATTGAA
 V A H T Y E L T S S S Y I Q N D P I T N P G N F L P R I S P Y Q L R V K C A A E
 580 570 560 550
 2770 2780 2790 2800 2810 2820 2830 2840 2850 2860 2870 2880
 AACAAATAAGTAAGCTTGAAGTCTGAGTAAATTAGCATTATAAGATTATCATCTAAATTAGCTGTTGCAATGCTGAACAGGTTCACCAATATGCAA
 V I F Y A Q T D T F N A N N L N D D L N A T T L A H V P K V G I H L S P Q A K P
 540 530 520 510
 2890 2900 2910 2920 2930 2940 2950 2960 2970 2980 2990 3000
 TGTTATAAGATGGAAAAGCTCTGATGAAGTATTGACTTTCAATTGAGTCCATTAGCTGTTGCAATAGTCCATTAGTATTGATACAGTATCATTAA
 T Y S P F A G A H L I Q S K E I Q T A I T W E T N S V T D N F K E S D G R L I
 500 490 480 470
 3010 3020 3030 3040 3050 3060 3070 3080 3090 3100 3110 3120
 ATTATTAGTCATAGTAAACAGCAGTCTGAGACTCGAGTTCCCTGACCTGCTGGCAACTAACAGGTTGAGTACACCCACTGGCTAGGTACACCAG
 N N T M F D A T R Y Q S N G T P C S L T S N G V A S P V G T Y V A P I A
 460 450 440 430
 3130 3140 3150 3160 3170 3180 3190 3200 3210 3220 3230 3240
 TTTTTAAATTAAACCCAATTGGTTATAATTATATTCAACAAATTCTCTGGAGGTCCTACATCTAAATTCTACATGAGATTGAAACCTGGCGTAA
 K K I L G L Q P K Y N E V I E G P P G N C R I E N V H S Q L C E Y G P N D A T
 420 410 400 390
 3250 3260 3270 3280 3290 3300 3310 3320 3330 3340 3350 3360
 ACCGTTCTGTTGAAGAACAGTCGAAATAATTCTTAATGGATACCGAACTGATGACGTTGACATGATTGTTGACTGTTATCGCGTAAACCTTAA
 G N Q Q L V L A F Y N K L P Y P I G F Q H R P V H N P V S K D G Y F E K V Y D E
 380 370 360 350
 3370 3380 3390 3400 3410 3420 3430 3440 3450 3460 3470 3480
 AAAAGAGATTTCATCACGCTTGTCCAAGTTCATCTATGGATGATACAACCACTGGTTCTAGCTGAAATTCTTAGGAGCTACGTCATAACCTGT
 F L N E D R I A G L E D I S S V V M P E N A Q F E K P R V D I G Q T T K I N L G T
 340 330 320 310
 3490 3500 3510 3520 3530 3540 3550 3560 3570 3580 3590 3600
 AGCATGAAGAATAATTGATTGTTAGAGTAGCAAGACCACTTAACTGAACTATTGTTGAAAGAACATACGGGGGAAATGCTGTTACCGCTACATT
 A H L I F Q N Q N L T A L G T N S S N T Q F A I R P T F A T V R V N C E E I C S
 300 290 280 270
 3610 3620 3630 3640 3650 3660 3670 3680 3690 3700 3710 3720
 TCCAGGTGTAATTGAGCAATTCTTGGATTCTATACAAGAACGGTCTGGCGACTGTTAACCTGGCCAAACTGTAATAACCGAAGTTGTCAC
 G P P L Q A F E N P N M Y L F P R D V P L R G L S T I V F N D V Y T V D N V T R
 260 250 240 230
 3730 3740 3750 3760 3770 3780 3790 3800 3810 3820 3830 3840
 TGTTATTGGAAATTACAACATTGCAAGACCACTAGTTAAACATGGATGACTCTTGTAAAGATAATGAAATTGTTGCGGAGTAGTAATTGG
 T I P I V V N A L G Y T L F R H V K N F I I F N N H P T T I P R P I P I I A Q G
 220 210 200 190
 3850 3860 3870 3880 3890 3900 3910 3920 3930 3940 3950 3960
 GCCTCTTCACCTGCAATATCTGACCTGCTCCCGTAATGATGACCTGCTTGAACGTTTACCACTACTTTAGGAGGACGACCA
 G G E G A I D A G A G P L S T G A M R S R K G G S S K P P R G G G R K G K G K
 180 170 VP4 160 150
 3970 3980 3990 4000 4010 4020 4030 4040 4050 4060 4070 4080
 GTCTTGTTGTCGGGGTTATGGAGGAGTTAAGGAGCACTCATGGACAGAACATCCTGTCAGGTGGAGTACTATCAGGATTACTACAG
 D Q T K D P T I P P T L P A S M S S S D G E D P P T S D V P N S V P E P D P D T
 140 VP3 130 120 110

4090 4100 4110 4120 4130 4140 4150 4160 4170 4180 4190 4200
 GGAACGTCAGGGTGCCTGCTAAATCAAGTAGTCGGATGATGTTCTGTAGTAGAAGGAAGTTGATCTACGTCGAAGTCGCAATGTCCTCCGGATTACCTCAGGTAGCTCGCTAA
 S R G P Q A A L D T T P H H E Q T T S P L Q D V D F D D I D E P N G E P L E D L
 100 90 80 70
 4210 4220 4230 4240 4250 4260 4270 4280 4290 4300 4310 4320
 TATATCATCTGCAAGTTGGAGTTGGAGGTCGACTCTCTCTCTGACAGGAGGGTCTGGAAATAGGATGATCAATTGGTAAACCTCGACGTAACGAGCTAAATGCCATTGTTAAC
 I D D A T P T A P A V E E E E A R P P E P I P H D I P L G R R V R A L Q W Q E V
 60 50 /acceptor site alternative splicing II
 4330 4340 4350 4360 4370 4380 4390 4400 4410 4420 4430 4440
 AGCGTACCGCGGTTGACCTCATTTAAGTATCCAAATTACAGTTCTGTGGATGAGGTTCTAGTGGGGTGTCAATTACACTGAAAACAAAATAAAAGTATCAGTAACGAGCTTT
 A Y R R Q G E N L Y E W N D R E Q P H P K L P P T M V s f v f y f t d t s v n k
 20 10 * Y E R K
 alternative splicing II donor site √
 4450 4460 4470 4480 4490 4500 4510 4520 4530 4540 4550 4560
 TAGGTTGAGGATATACAGGTCACCAATTACGTTCTCAATTACAGTTAGCTTTAAAGTCGTATTTCACAAATGCTGGTATTCTGGTACCGGATTCTGGTATTCCTCAGAAA
 I n n l i y l d v l
 K P Q Q P Y V P G G I V D E I V R K A K L G Y K G V F A Q W T G S K Q Y E E S F
 260 250 240 230
 4570 4580 4590 4600 4610 4620 4630 4640 4650 4660 4670 4680
 GATGATTGCTCGTGTCCAATTATATACTTACGTTGAATCTCTCAGTAAACATCTGATTTCAGTGGCAAAAGTCTCGTACCCAAAGATCGTGTCTAGAATT
 R H I A E D D L K Y V E T K F D E E T F Y Q S K A R H L L D E Y G L D H E K A I
 220 210 200 190 PLA2
 4690 4700 4710 4720 4730 4740 4750 4760 4770 4780 4790 4800
 GATCAGTTCTGATCGTCAGGTTAGGATCTATTGGATTACCGAGCCCCACGAAATCAGTGCCAGGTAACACTGCACCTTGCTACTCGTGTGATCAATTAA
 Q D T E S R A P K P D I P N G P G V F D T G P L V A G K S S T H D I L K K I G I
 180 170 Ca² 160 150
 4810 4820 4830 4840 4850 4860 4870 4880 4890 4900 4910 4920
 CGATTCCTGCGAGGGCTGCTCTCTGCTGAGTGGCCGACAGTTCAAGAGTTCCAGCTGTGCTGCGTACCTGAGTAGGCCCTCCAAAAGTCTGTTGGTAG
 G I G A L A A G G A A T A A V T G L T G A T G A A A S G A T A G G G L L G T T E T
 140 130 120 110
 4930 4940 4950 4960 4970 4980 4990 5000 5010 5020 5030 5040
 CTGCTTCGCCGAGTTCAAGTGCTTTCTCAACTCCAGTAGAGAAACTTGTTCAGGTTGGCTCTCCACCTGCGTTGATTTGCTTCGGGATATAATCGGAGTTCTCTG
 A A E G L E L A T E E V G T S F S T E A E A A G E G A N I E D E A I D I R T E E
 100 90 80 70
 5050 5060 5070 5080 5090 5100 5110 5120 5130 5140 5150 5160
 GCCCTGCCGCTCTGGAAACTAGTTGGCAAGAGGTTGGTCAATTGTAGGTAGTCGTCAGGGTGACTACGATTCCACCATCTCCAAATCTATTACGAAGTTAGGTGAT
 P G A A E Q F S T E R L P E P E N T P L R D Y P H S R N W W R G F R N R L K P T
 60 50 40 30
 5170 5180 5190 5200 5210 5220 5230 5240 5250 5260 5270 5280
 TTGCTAAAGGTAATTATTAGTACCTAAACTGAAATCACGCAACGTGAGAGATCTGAAAGACGCCAGACATTAAACTACGATGCTAAAACCTCCGCTTATACCTCTGACGTCAGCG
 N A L P L N N T G L S F D R L T L D T F V G S M *
 20 10 Start transcription
 ITR
 | 5290 5300 5310 5320 5330 5340 5350 5360 5370 5380 5390 5400
 AAATAGAAGAAGTGCCTGGGTGCTGCTCTTATACCTCGGATATAACAACAACTCCCACCAACCTGACCTGAACCTAAAGTTCAAGGTCAGGTTGGCTGGGGAG
 TTGTTGTTTATACCGAAGGATATA

Figure S1: Annotated sequence of the AdDNV strain isolated in 1977 by Meynardier et al. Translated sequence above nucleotide sequence is from sense mRNA, whereas the translation below the translation is from the antisense mRNA. Both the nucleic acid and the proteins are numbered; however, the NS2 sequence and numbering is in italics since it overlaps NS1. Inverted terminal repeats are indicated by ITR, transcription starts, stops and stop codons by an asterisk. Slashes indicate splicing sites. Rolling-circle replication motifs (I and II), Walker motifs (A, B, C), phospholipase A2 (PLA2) and its Ca² (Ca²) motifs are highlighted in gray. The different splicing combinations in the VP ORFs are distinguished by Ia, Ib and II. Tata-boxes are underscored.

Monoisotopic mass of neutral peptide Mr(calc): 1079.54

Ions Score: 61 **Expect:** 1.4e-005

Matches (Bold Red): 16/92 fragment ions using 28 most intense peaks

#	a	a ⁺⁺	a*	a* ⁺⁺	b	b ⁺⁺	b*	b* ⁺⁺	Seq.	y	y ⁺⁺	y*	y* ⁺⁺	#
1	60.04	30.53			88.04	44.52			S					10
2	117.07	59.04			145.06	73.03			G	993.52	497.26	976.49	488.75	9
3	218.11	109.56			246.11	123.56			T	936.49	468.75	919.47	460.24	8
4	404.19	202.60			432.19	216.60			W	835.45	418.23	818.42	409.71	7
5	532.25	266.63	515.22	258.12	560.25	280.63	543.22	272.11	Q	649.37	325.19	632.34	316.67	6
6	603.29	302.15	586.26	293.63	631.28	316.15	614.26	307.63	A	521.31	261.16	504.28	252.64	5
7	750.36	375.68	733.33	367.17	778.35	389.68	761.33	381.17	F	450.27	225.64	433.24	217.13	4
8	849.43	425.22	832.40	416.70	877.42	439.21	860.39	430.70	V	303.20	152.10	286.18	143.59	3
9	906.45	453.73	889.42	445.21	934.44	467.72	917.42	459.21	G	204.13	102.57	187.11	94.06	2
10									K	147.11	74.06	130.09	65.55	1

Monoisotopic mass of neutral peptide Mr(calc): 1751.95

Ions Score: 87 **Expect:** 2.1e-008

Matches (Bold Red): 24/132 fragment ions using 38 most intense peaks

#	a	a ⁺⁺	a*	a* ⁺⁺	b	b ⁺⁺	b*	b* ⁺⁺	Seq.	y	y ⁺⁺	y*	y* ⁺⁺	#
1	86.10	43.55			114.09	57.55			L					16
2	199.18	100.09			227.18	114.09			I	1639.87	820.44	1622.84	811.92	15
3	328.22	164.62			356.22	178.61			E	1526.78	763.90	1509.76	755.38	14
4	443.25	222.13			471.24	236.13			D	1397.74	699.37	1380.72	690.86	13
5	542.32	271.66			570.31	285.66			V	1282.72	641.86	1265.69	633.35	12
6	655.40	328.20			683.40	342.20			I	1183.65	592.33	1166.62	583.81	11
7	712.42	356.72			740.42	370.71			G	1070.56	535.79	1053.54	527.27	10
8	769.45	385.23			797.44	399.22			G	1013.54	507.27	996.51	498.76	9
9	866.50	433.75			894.49	447.75			P	956.52	478.76	939.49	470.25	8
10	965.57	483.29			993.56	497.28			V	859.47	430.24	842.44	421.72	7
11	1128.63	564.82			1156.62	578.82			Y	760.40	380.70	743.37	372.19	6
12	1225.68	613.34			1253.68	627.34			P	597.34	299.17	580.31	290.66	5
13	1353.74	677.37	1336.71	668.86	1381.74	691.37	1364.71	682.86	Q	500.28	250.64	483.26	242.13	4
14	1481.80	741.40	1464.77	732.89	1509.79	755.40	1492.77	746.89	Q	372.22	186.62	355.20	178.10	3
15	1578.85	789.93	1561.83	781.42	1606.85	803.93	1589.82	795.41	P	244.17	122.59	227.14	114.07	2
16									K	147.11	74.06	130.09	65.55	1

Figure S2: MS results.

Mass spectrometry of SGTWQAFVGK and LIEDDVIGGPVYPQQPK peptides that are present in the 816-aa protein but not in the potential 783 and 708 aa structural proteins (in the introns of the latter).

NS1	576 amino acids		VP (ORF-A)	597 amino acids	
	40%, +56% (559, 4%), 7e-105	AdDNV		35%, +50% (604, 10%), 1e-70	AdDNV
	28%, +48% (580, 8%), 6e-57	PcDNV		31%, +46% (587, 8%), 1e-47	PcDNV
	28%, +47% (587, 10%), 5e-52	CpDNV		35%, +49% (406, 7%), 2e-43	PfDNV
	31%, +47% (488, 11%), 4e-34	DsDNV		30%, +46% (434, 10%), 1e-39	JcDNV
	26%, +43% (555, 14%), 6e-31	PfDNV		31%, +46% (478, 9%), 7e-39	MpDNV
	25%, +43% (520, 10%), 2e-25	BgDNV		28%, +41% (637, 12%), 1e-36	CpDNV
	25%, +41% (443, 13%), 1e-18	DpIDNV		28%, +43% (602, 14%), 4e-32	DpIDNV
	30%, +49% (141, 5%), 1e-6	DpDNV			BgDNV
	30%, +49% (207, 9%), 3e-17	MpDNV			
NS2	286 amino acids		VP (ORF-B)	268 amino acids	
	34%, +48% (221, 4%), 6e-15	AdDNV		33%, +44% (233, 14%), 1e-14	AdDNV
	30%, +43% (224, 5%), 1e-9	MIDNV		33%, +50% (101, 14%), 1e-4	PfDNV
	26%, +41% (243, 3%), 2e-9	PfDNV		38%, +51% (123, 14%), 5e-04	DpDNV
	22%, +39% (252, 9%), 0.078	CpDNV		33%, +51% (104, 5%), 0.001	PcDNV
		BgDNV		30%, +46% (106, 15%), 0.001	DsDNV
				30%, +53% (103, 17%), 0.005	BmDNV
					BgDNV

Figure S3: Similarities of AdDNV ORF sequences with those of other densoviruses.

The lines represent the location of the homologous domains in the other densoviruses. Above each line, the identity %, positives %, (length of domain in amino acids, gap %), likelihood of similarity by chance, is noted. Lower lines indicate increased likelihoods of similarity by chance as determined by BLAST (NCBI). Only one DNV was included from the Densovirus and Iteravirus genera since values of their members were very similar. Many hits were obtained for NS1 and ORF-B because they contained enzyme motifs (NTPase and vPLA2, respectively). AdDNV NS3 did not yield homologous proteins with BLAST and relatively few for NS2. The best E-values were obtained for NS1 and ORF-A. Also, the largest coverage of ORF-A was obtained with densoviruses that also had a split VP cassette (PcDNV, PfDNV, DpIDNV, BgDNV).

Article 6: A Novel Ambisense Densovirus, *Acheta domesticus* Mini Ambidensovirus, from Crickets

Hanh T. Pham, Qian Yu, Max Bergoin, Peter Tijssen

INRS-Institut Armand-Frappier, Laval, QC, Canada

Contribution of authors

Hanh Pham was mainly responsible for this work including virus purification from infected crickets, DNA extraction, cloning, and result analysis. Qian Yu has contributed in the cloning steps, sequences and results analysis. Peter Tijssen and Max Bergoin planned and supervised the project. Hanh Pham and Peter Tijssen prepared the manuscripts.

Abstract

The genome structure of *Acheta domesticus* mini ambidensovirus, isolated from crickets, resembled that of ambisense densoviruses from *Lepidoptera* but was 20% smaller. It had the highest (<25%) protein sequence identity with the nonstructural protein 1 (NS1) of Iteravirus and VP of *Densovirus** members (both with 25% coverage) and smaller (0.2- versus 0.55-kb) Y-shaped inverted terminal repeats.

Résumé

Le génome du virus « *Acheta domesticus mini ambidensovirus* » a été isolé à partir de criquets, est apparenté au densovirus ambisens de *Lepidoptera*, mais il était 20% plus petit. Il avait la plus haute identité de séquence protéique (<25%) avec la protéine non structurale NS1 des Iteravirus et les protéines VP des membres du genre *Densovirus** (les deux avec une couverture de 25%) et il avait des répétitions terminales inversées en Y plus petites (0.2kb comparativement à 0.55kb).

Results

The cricket industry has been devastated worldwide recently by the *Acheta domesticus* densovirus (AdDNV) (Liu *et al.*, 2011, Pham *et al.*, 2013b, Szelei *et al.*, 2011, Weissman *et al.*, 2012). We also observed several, thus far unknown, viruses such as volvoviruses, which have circular, single-stranded DNA (ssDNA) genomes (Pham *et al.*, 2013c), and a new densovirus (parvovirus).

Two genera of insect parvoviruses, named densoviruses (Tijssen *et al.*, 2011), are particularly relevant for this new densovirus. The *Densovirus** genus contains ambisense densoviruses from *Lepidoptera*, with genomes of 6 kb, Y-shaped inverted terminal repeats (ITRs) of about 0.55 kb, and sequence identities of about 85% (El-Far *et al.*, 2012, Fédière *et al.*, 2004, Huynh *et al.*, 2012, Pham *et al.*, 2013d, Tijssen *et al.*, 2003). The *Iteravirus** genus contains monosense densoviruses, also from *Lepidoptera*, with 5-kb genomes, J-shaped 0.25-kb inverted terminal repeats (ITRs), and about 75% sequence identities (Fédière *et al.*, 2002, Li *et al.*, 2001, Yu *et al.*, 2012a, Yu *et al.*, 2012b).

A new virus with morphology and size similar to densoviruses was detected in some cricket samples from the United States. Virus was purified and DNA extracted as described previously (Pham *et al.*, 2013c). Digestion of viral DNA with *Eco*RI yielded 2 bands of about 700 bp and 4,200 bp on agarose gels. DNA was blunt ended with T4 DNA polymerase and a large Klenow fragment in the presence of dNTPs at room temperature (RT), ligated into the *Eco*RV site of pBlueskript KS(-), and transformed into SURE cells. DNA of clones with expected sizes were subcloned. Digestion with *Eco*NI within the terminal hairpins yielded clear reads of ITR sequences. Several complete clones were sequenced in both directions by use of Sanger's primer-walking method as described previously (Tijssen *et al.*, 2003). Contigs were assembled by use of the CAP3 program (<http://pbil.univ-lyon1.fr/cap3.php/>) (Huang *et al.*, 1999).

Surprisingly, the genome structure and gene organization of this virus strongly resembled those of ambisense densoviruses from the *Densovirus** genus (El-Far *et al.*, 2012, Fédière *et al.*, 2004, Huynh *et al.*, 2012, Pham *et al.*, 2013d, Tijssen *et al.*, 2003), but the genome sequence was only 4,945 nucleotides (nt) long, instead of about 6,035 nt, and lacked nucleotide sequence identity (best E value of 0.017, with a query coverage of 1%). Protein sequence identities were for the major nonstructural protein 1 (NS1) closest to *Iteravirus** members and, oddly, for the structural proteins (VP) closest to *Densovirus** members (both at best 25% identity for 25% coverage [or higher for shorter coverage]).

ITRs of AdMADV were smaller than those of *Densovirus** members (199 versus about 545 nt) and Y-shaped, with a 113-nt hairpin. The 45-nt-long stem contained two side arms in the middle, nt 46 to 68, that occurred in two sequence orientations (flip/flop). It had a high GC content (63%) and contained inboard TATA boxes, at 193 to 199 for the NS cassette and at 4747 to 4753 for the VP cassette. This structure is identical to that of *Densovirus** ITRs.

The NS cassette consisted, as for *Densovirus** members, of NS3, followed by NS1 and an overlapping NS2. Splicing, as for *Densovirus**, would remove the NS3 open reading frame (ORF) and allow expression of NS1 and NS2 by leaky scanning. As for Densovirus, the putative splice acceptor site was located just upstream of the initiation codon of NS1 (1172-CAG/aATG_{NS1}...N₁₉...ATG_{NS2}) (in GmDNV, 1395-CAG/ATG_{NS1}...N₄...ATG_{NS2}). As for members of the *Densovirus** genus, the VP on the complementary strand also contained the phospholipase A2 motif (4,590 to 4,680 nt) (Zádori *et al.*, 2001) and the stop codons of NS1 and VP were neighbors (2661-TAG/AAT-2666), suggesting a small overlap of their transcripts, as for GmDNV.

***Note:** Since publication of this manuscript, the name of the Densovirus and Iteravirus genera have been changed to Ambidensovirus and Iteradensovirus, respectively (Cotmore *et al.*, 2014a).

Nucleotide sequence accession number. The GenBank accession number of AdMADV is KF275669.

ACKNOWLEDGMENTS

This work was supported by the Natural Sciences and Engineering Research Council of Canada to P.T.; H.T.P. and Q.Y. acknowledge tuition fee waivers at INRS and scholarships from INRS and the People's Republic of China.

In progress:

Novel densovirus from crickets include a segmented densovirus

Hanh Thi Pham[§], Qian Yu[§], and Peter Tijssen

INRS-Institut Armand-Frappier, Université du Québec, 531 boul. des Prairies, Laval, Québec, Canada, H7V1B7

§ H.T.P. and Q.Y. contributed equally to this work

Abstract

Densoviruses are members of the *Parvoviridae* family infecting invertebrates. Since 2009, a cricket densovirus, AdDNV, caused a severe epidemic in the \$ 600-million cricket industry in North-America. During the last 18 months, we also received samples of diseased crickets from North America, Europe and Japan that were negative for AdDNV. Electron microscopy demonstrated that these samples contained densovirus-like particles. A SISPA method, which has been introduced previously in this thesis, was used to detect the pathogens (Yu *et al.*, 2014b). Sequencing identified those amplicons with identities to known viruses and allowed by primer extensions to obtain the complete genomes and by cloning in suitable vectors to obtain infectious clones. Sequencing revealed that some were novel circoviruses (Pham *et al.*, 2013c) and a novel ambisense densovirus (Pham *et al.*, 2013f). However, the greatest surprise was a densovirus with a segmented genome (*Acheta domesticus* Segmented Densovirus, AdSDV). NS- and VP- coding sequences are in separate segments of about 3.3 kb. Both NS- and VP- segments are flanked by identical hairpin terminal sequences but lack ITRs. The two ORFs (NS1 and NS2) within the NS-segment have 38% and 40% amino acid identity to the mosquito brevidensovirus CppDV and AalDV, respectively. The upstream ORF in the VP segment also has a relatively high identity with these mosquito densoviruses whereas the downstream ORF has a relatively high identity with the cricket AdDNV of the ambidensoviruses and arose

through recombination between a mosquito brevidensovirus and AdDNV. This downstream ORF with the AdDNV-VP1up-like phospholipase A2 domain could be spliced to the C-terminus of the major capsid protein. Here, we will present a hypothesis on the origin of this first bipartite genome parvovirus that arose from mosquitoes to infect crickets. Currently, the research of capsid structure (X-ray crystallography) and transcription strategy of AdSDV is underway. Meanwhile, in a similar vein the VP gene of AdDNV probably underwent recombination with an aphrodisiac iridescent virus (iridovirus) of crickets. These findings provide new insights in cricket densovirus evolution and also necessitate a revision of the definition and taxonomy of parvoviruses.

Results

Cloning of the complete genome of AdSDV

Virus was partially purified from AdDNV-negative cricket samples from different farms in Canada and USA by the method described for GmDNV (Tijssen *et al.*, 2003) and visualized by electron microscopy (Figure 5.7B). A DNA genome of over 3 kb was suggested by agarose gel electroporesis (Figure 5.7A). Viral DNA was then analyzed by several restriction enzymes, *Hind*III, *Spel* and *Xba*I. Digesting the viral DNA with *Hind*III or *Spel* yielded two smaller bands and the native DNA band. While, digesting the viral DNA with *Xba*I only gave rise to two smaller fragments of around 1.3 kb and 2.0 kb (Figure 5.7C). The digestion profiles indicated the undigested native viral DNA may be not one but a combination of two similar sized segments of the viral genome. *Hind*III and *Spel* are specific enzymes for each genome segment, and the *Xba*I sites exist in both genome segments at almost the same position. The viral DNA was then blunt-ended by Klenow large-fragment, digested with *Hind*III or *Spel*, and cloned into a pBluescript SKII(-) vector. Sequence analysis by Sanger's method (Yu *et al.*, 2012a) revealed that NS- and VP- coding sequences were in fact separate segments of 3313 nts and 3332 nts, respectively. Both NS- and VP- segments are flanked by identical hairpin terminal sequences but lack ITRs. Similar to other brevidensoviruses, the telomeres at both ends are dissimilar. The left hairpin was 208

nts with 43% GC content, while the right hairpin was 220 nts with 48% GC content (Figure 5.8A). However, the left hairpins in the two segments were identical as were the two right hairpins.

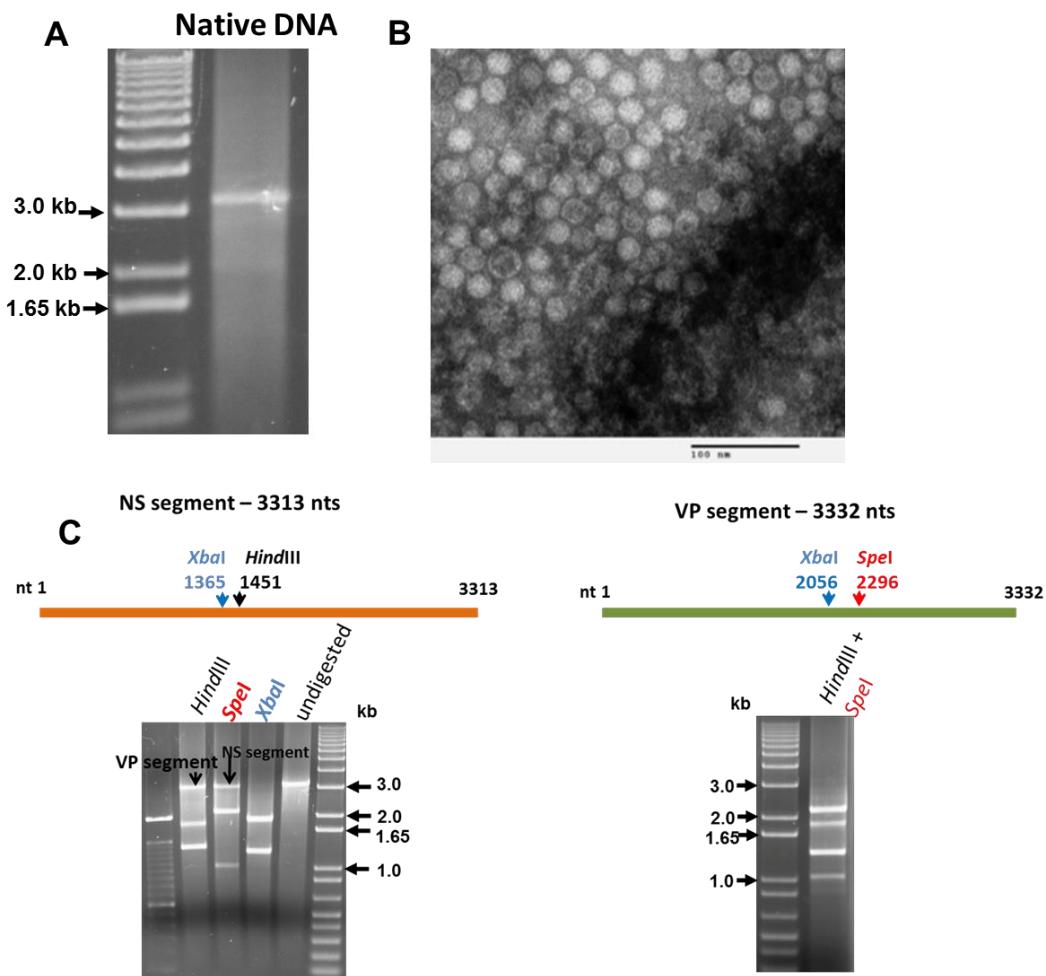


Figure 5.7 Genome analysis of AdSDV.
A) The native viral DNA by agarose gel electrophoresis showed only one band at around 3.3 kb.
B) Electron microscopy of virus partical. (Bar: 100nm)
C) Restriction analysis of viral DNA. Left gel: 1st lane is HindIII digestion; 2nd lane is SpeI digestion; 3rd lane is XbaI digestion (two doublets of 2056 and 1948 nts and of 1365 and 1276 nts); 4th lane is the undigested viral DNA. Right gel: HindIII and SpeI double digestion. Marker: 1 kb DNA ladder.

Genome organization of AdSDV

Using the NCBI ORF-Finder program, two large ORFs for each segment genome were identified (Figure 5.8B). The large ORF1 on the NS-segment had the capacity to

code for brevidensovirus-like NS1 with a putative a.a. number of 795. ORF2 corresponds to NS2 with a capacity of 436 a.a. NS1 had about 38% a.a. identity to the NS1 protein of mosquito brevidensoviruses like CppDV (Zhai *et al.*, 2008). NS2 had about 40% a.a. homology to the NS2 protein of mosquito brevidensoviruses like AaIDV (Chen *et al.*, 2004). The NS1 also contained the highly conserved RCR motif and NTPase motif with sequence identity and location similar to that of other densoviruses as well as vertebrate parvoviruses. For the VP-segment, a large ORF1 at the 5'-end with a potential coding region of a 377 a.a. had about 39% a.a. identity to mosquito brevidensoviruses. However, the large ORF2 at the 3'-end with a capacity of 346 a.a. contained the conserved PLA2 motif and had 40% a.a. identity to the upstream ORF of VP1 of cricket ambidensovirus AdDNV.

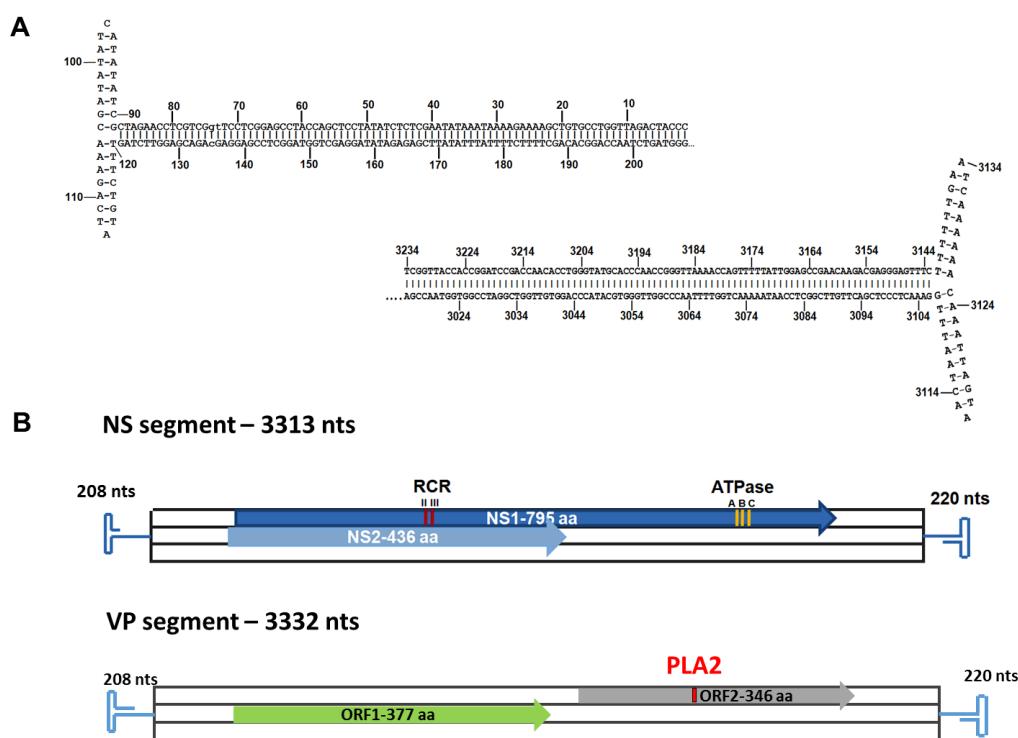


Figure 5.8 Telomeres and genome organization of AdSDV.
A) The telomeres of AdSDV. Left hairpin of 208 nts (upper) and right hairpin of 220 nts (lower).
B) Genome organization of AdSDV. Both NS and VP segments contain the same telomeres on left and right ends. NS cassette includes two large ORFs, NS1 and NS2. The NS1 also contains the conserved RCR and ATPase motifs. The VP cassette contained two large ORFs without overlaps. The ORF1 may response for the major capsid. The ORF2 at C-terminus of VP contains the conserved PLA2 motif.

Recombinant cricket aphrodisiac iridescent virus

In 2014, we received some virus sample purified from infected cricket *Acheta domesticus* and started to work on it as a collaboration with Prof. Drion G. Boucias of the University of Florida. All the infected crickets were sick and dying and filled with some milk-like fluid (Figure 5.9A), which is a distinct symptom from AdDNV-infected crickets. Sets of primers targeting the NS and VP gene of AdDNV were used to detect the infected cricket samples collected in different years in Prof. Boucias's lab. Compared to earlier samples, more recent isolates 2013 and 2014 could only be amplified with primers targeting the AdDNV VP gene (Figure 5.9C). Alternative NS primer sets also failed to amplify AdDNV NS (results not shown). Most interestingly, electron microscopic visualization of purified virions from the 2014 isolates revealed an abundance of particles of about 150 nm in diameter but not of 25-nm densovirus-like particles. Given the EM results, and from descriptions in a recent publication (Adamo *et al.*, 2014), the virus might be identified as cricket aphrodisiac iridescent virus (iridovirus). However, since only the AdDNV VP gene has been also detected from this sample, a mixed infection, or, most probably, a horizontal transmission of cricket densovirus VP gene to iridovirus may be possible. More analysis should be used for this new cricket virus next. If it is indeed a recombination of two viruses, this will be another example of about the cricket densovirus recombination and evolution.

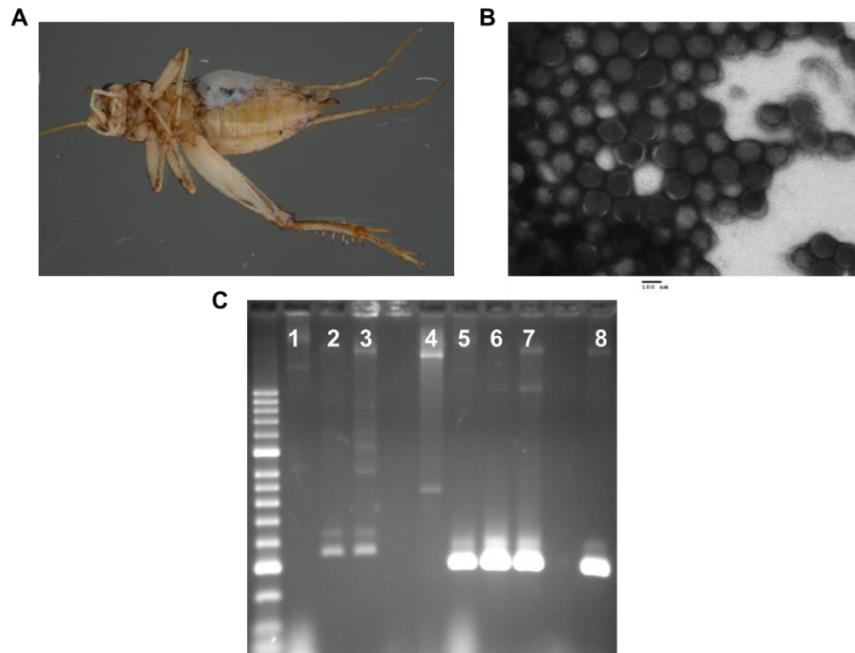


Figure 5.9 The new, possibly recombinant, cricket aphrodisiac iridescent virus.
A) Infected cricket exuding a milk-like fluid. **B)** Virus particles shown by electron microscopy. (Scale bar: 100 nm) **C)** PCR amplification of the virus from infected crickets collected in different years with AdDNV NS and VP gene specific primers. Lane 1-4: AdDNV NS. Lane 5-8: AdDNV VP. Lane 1, 5: virus from infected crickets in 2013; Lane 2, 6: virus from infected crickets in 2010; Lane 3, 7: virus from infected crickets in 2011; Lane 4, 8: virus from infected crickets in 2014.

Disscussion

The AdDNV epidemic appears to have declined since the end of 2012. However, some cricket farms in US and Canada still struggle with unknown diseases. The new cricket densovirus we reported here was first observed from an AdDNV-negative cricket sample from a cricket farm in Ontario (Canada). Then, detection of abundant cricket samples from different farms in the USA showed that this new virus was related to a recent epidemic. Electron microscopic visualization of purified virions from these samples revealed the abundance of small particles about 22-23 nm in diameter. Restriction enzymes analysis, genome cloning and sequencing were used to identify this new virus. The greatest surprise was the NS and VP genes are in separated segments and flanked by similar hairpins that may be important for replication of the segments. NS proteins and the N-terminus of the VP protein share highest a.a.

similarity with those of the *Brevidensovirus* genus. Moreover, the ORFs on the NS segment are organized as for the brevidensoviruses and follow probably a similar expression strategy. The viral genome contains all the critical characteristics of a typical densovirus (RCR, NTPase and PLA2 motifs). However, the structural proteins are C-terminal extended by an *Ambidensovirus*-like sequence and the PLA2 motif is, in contrast to the vertebrate parvoviruses and densoviruses, located at the C-terminus. We provide here a hypothesis on the origin of this segmented cricket densovirus (AdSDV) which may explain how a brevidensovirus that normally infects only mosquitoes has adapted itself through recombination with AdDNV to crickets (Figure 5.9). The ancestral brevidensovirus-like virus (no PLA2 motif) adapted to infect cricket via a recombination with the cricket ambidensovirus AdDNV. The N-terminus of AdDNV VP was inserted into C-terminus of VP of the ancestral brevidensovirus-like virus. This recombination enlarged the ancestral virus genome to that of a large protovirus, which included the complete viral genome of the ancestral virus and the N-terminus part of AdDNV VP (with PLA2 motif). The PLA2 domain resides normally within the N-terminus region of VP1up of all known parvoviruses known to have this activity and is required for successful infection (Zadori et al., 2001). Brevidensovirus, hepandensovirus and Aleutian mink disease virus (AMDV) are the only parvoviruses that do not have PLA2 motifs. This may suggest that the protovirus obtained a phospholipase A2 activity in order to enable it to infect crickets. However, the brevidensovirus capsid might have been too small for the enlarged genome and enlarged protein with the PLA2 domain. Thus, a genome segmentation might have been necessary for successful assembly. Either a deletion of VP or a deletion of NS may happen within the protovirus. After the deletions, the segmented genomes contained the identical telomere, which may be required for the successful packaging and rolling-circle replication. The two ORFs for VP arose through recombination between a mosquito brevidensovirus and AdDNV. The downstream ORF with AdDNV-VP1up-like phospholipase A2 could be spliced to the C-terminus of the major capsid protein. Since its genetic characteristics are interestingly unique among all the densoviruses and vertebrate parvoviruses, a special taxonomic classification may be needed for this virus.

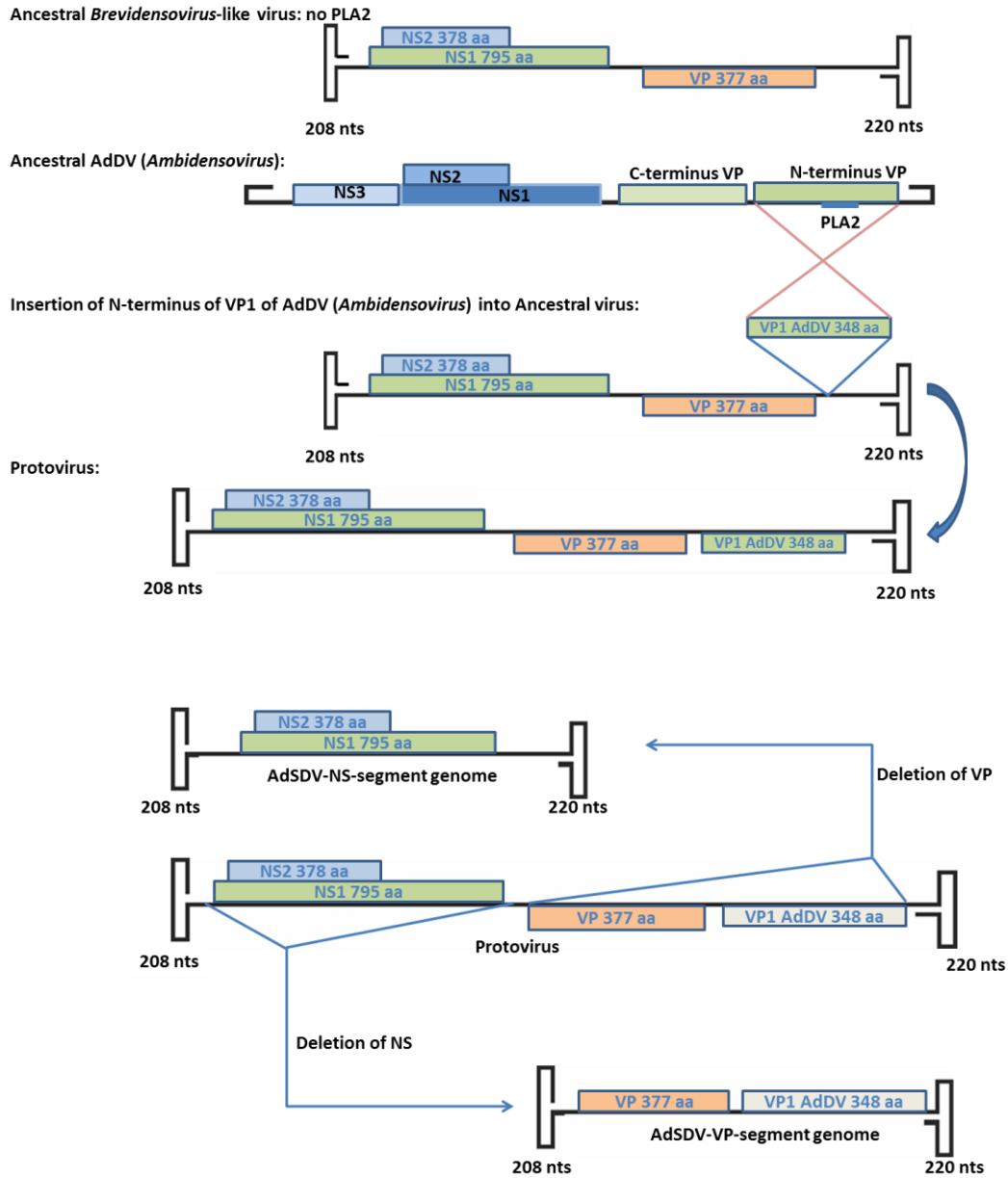


Figure 5.9 Hypothesis on the origin of AdSDV.

Chapter VI: DISCUSSION

1 The discovery of new iteradensoviruses

There were only three iteradensoviruses reported before 2010: BmDNV-1 from *Bombyx mori* (Nakagaki *et al.*, 1980), CeDNV from *Casphalia extranea* (Fédière *et al.*, 2002) and DpDNV from *Dendrolimus punctatus* (Wang *et al.*, 2005). Over the past three years, our lab was asked to test for unknown pathogens from different insect samples. A Sequence Independent Single Primer Amplification (SISPA) (Reyes *et al.*, 1991) method was used for the pathogen identification. This method allows nucleic acids of unknown sequence to be amplified with sequence-independent PCR methods, in which a primer-binding sequence is ligated to both ends of a cDNA fragment that can be PCR-amplified with a single primer (Allander *et al.*, 2001). We also incorporated a DNase and RNase treatment to remove the background nucleic acid of the hosts. PCR products were cloned and sequenced. The sequences obtained were subjected to homology search with BLAST (<http://www.ncbi.nlm.nih.gov/blast/>). Analysis of the results revealed high sequence identity with the members belonging to *Iteradensovirus* genus. This amplification and selection method was successfully used for identifying three new viruses from infected insects. Based on the partial genome sequence obtained from DNase&RNase-SISPA method, gene-specific primers were designed for whole genome amplification. Compared to the whole genome sequences of CeDNV, BmDNV-1 and DpDNV, the genomes of these three new viruses have 70-85% sequence identity, particularly in their typical J-shaped ITRs. Therefore, the three new viruses isolated from *Papilio polyxense* larvae, Monarch butterfly pupae and *Sibine fusca* larvae were later named as *Papilio polyxense* iteradensovirus (PpDV), *Danaus plexippus plexippus* iteradensovirus (DppIDV) and *Sibine fusca* iteradensovirus (SfDV).

This DNase&RNase-SISPA method is a useful tool for pathogen detection from insect samples. This method is suitable for screening of a large number of samples and also enables rapid sequence determination of high-titer viruses. However, even when the SISPA method was combined with a DNase and RNase treatment, the background material (e.g., host nucleic acid and environmental contaminants) were sometimes problematic. For samples of limited material and low viral titer, this

technique may potentially complicate identification of the target sequence(s) which is an important limitation and one of the greatest confounding factors still present in current pathogen discovery approaches (Firth *et al.*, 2013). Nowadays, deep sequencing (a.k.a High Throughput Sequencing) has revolutionized the field of pathogen discovery. Several DNA sequencing platforms have been developed. Deep sequencing is highly efficient, producing an enormous amount of information at low cost in a relatively short period of time (Quiñones-Mateu *et al.*, 2014). The development of this technique allows the rapid identification and characterization of novel viruses, and will also be fit for the discovery of novel parvoviruses, which are difficult to identify with classical technologies.

2 Cloning and sequencing of new iteradenovirus genomes

Cloning the full-length genomes of densoviruses is difficult due to the palindromic sequences at each end. These sequences form a hairpin-like structure in either “flip” or “flop” orientations as the palindromes are imperfect complements. The length and structure of the palindromes vary in different genera of the *Densovirinae* subfamily. For whole genome cloning of densoviruses, both the use of *Sure* bacteria as competent cells and culturing at 30°C instead of 37°C were found to be essential for amplifying the complete hairpins. To obtain infectious clones for all three new iteradenoviruses in this project, we developed a method for preparing Blunt-ended genomes. We found that for around 4 μ g purified viral DNA, the treatment for generating blunt ends combined 10 units of Klenow (LG Frag) polymerase and 12 units T4 DNA polymerase in the presence of 200 μ M of each dNTP yielded the best results for blunt-ending. Subsequent site-specific digestion near the middle of the viral genomes, was much more convenient to obtain separate clones that included the complete hairpins of both ends. To obtain the complete hairpin sequence by the Sanger method, two primers were used to read from both directions of the hairpins.

All five iteradenoviruses studied in this project have high identities in their genome organization. BmDNV contains the longest genome (5073 nts) among the five

viruses, due to the intergenic direct repeat of 45 nt between NS1 and VP coding regions (Li *et al.*, 2001). The differences in length between the members' genomes could be partly accounted for by differences in the lengths of the noncoding sequences between the ITRs and the viral genes. All the iteradensoviruses contained inverted terminal repeats (ITRs) including J-shaped hairpins at both ends of their genomes. Whole genome sequences had greater than 75% identities between one another. Figure 5.1A shows the phylogenetic tree for all the members of *Iteradensovirus* genus based on alignment of their whole genome sequences. Their monosense genomes contain three large ORFs, two of them on the left coding for the non-structural protein NS1 and NS2, and one large ORF on the right coding for the structure protein VP. Both NS and VP ORFs contain several domains that were conserved with other parvoviruses, such as the DNA-dependent ATPase domain in the NS1 cassette and the phospholipase A2 (pLA2) domain at the VP1 N-terminal extension (Figure 6.1B). The sequence identities of NS genes were higher than those of the VP genes (Figure 6.1B).

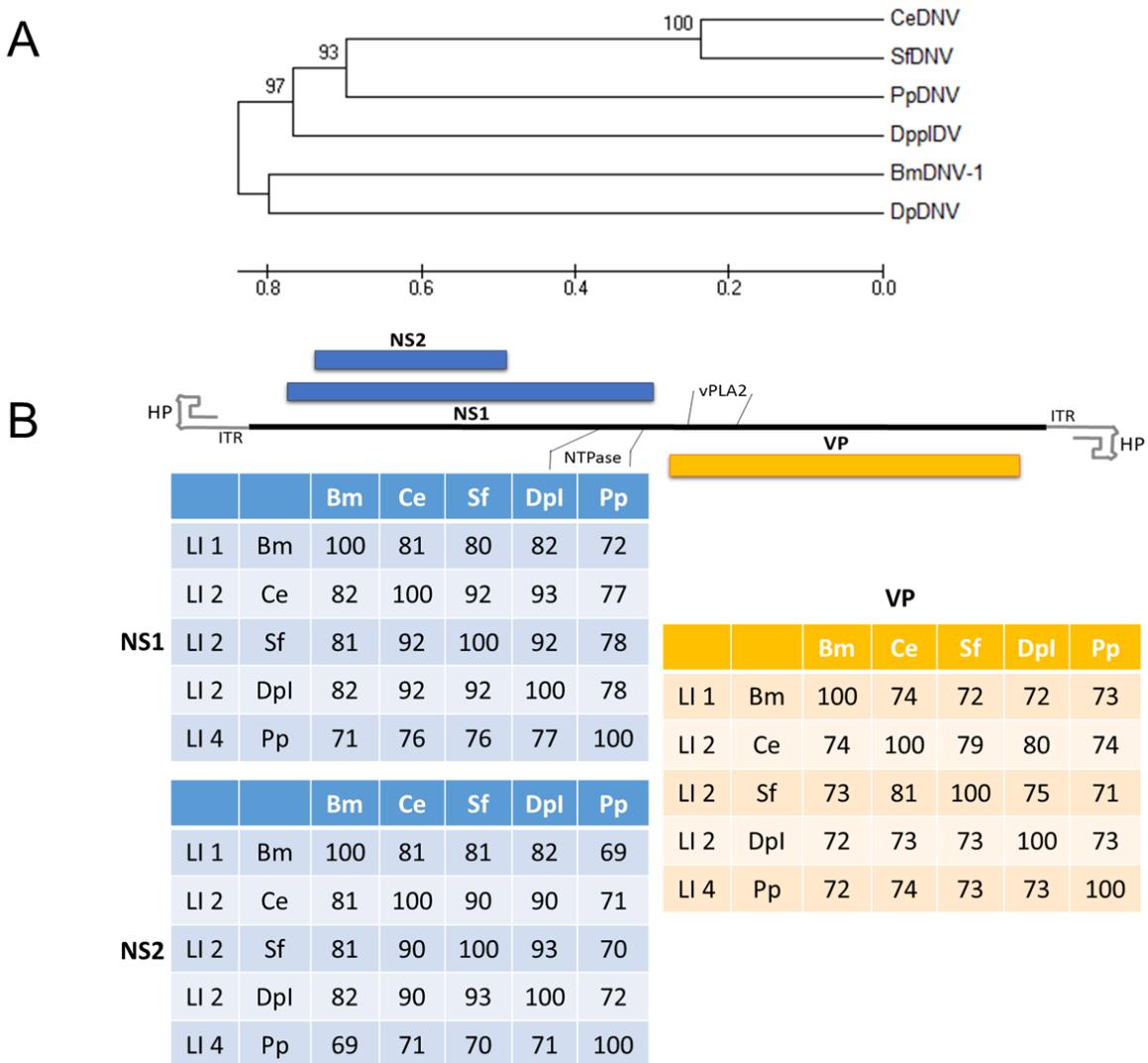


Figure 6.1 Genome comparison of iteradensoviruses.

A) Phylogenetic tree of the members in *Iteradensovirus* genus. Phylogenetic analysis based on the whole genome sequence of each virus (including the terminal sequences). The tree was built by Mega 6 software. Viruses including PpDNV, CeDNV, SfDNV, DppIDV, BmDNV and DpDNV. **B)** Three large ORFs of iteradensoviruses and comparison of their genome sequences. The upper schematic diagram show the NS1 and NS2 cassette on the left and the VP cassette on the right. Conserved domains of NS and VP were also marked (NTPase of NS and PLA2 of VP). The lower tables show sequence comparison of five iteradensoviruses, the numbers indicate Query coverage of nucleotide x sequence identity. LI1, LI2 and LI4 mean the viruses have been classified into different groups of *Iteradensovirus* genus (LI: Lepidopteran Iteradensovirus).

3 Expression strategy of iteradensoviruses

Despite the large number of densoviruses newly described in the last two decades, progress in understanding features of the densovirus life cycle, particularly strategies and regulation of genome expression as well as the viral proteins and their functions, has been slow and limited. The expression strategies have only been reported for certain species, such as the members of *Ambidensovirus* genus (GmDNV, JcDNV, PfDNV and BgDNV), the *Brevidensovirus* AaIDNV and the *Pestyldensovirus* PstDNV (Pham *et al.*, unpublished data). For the other genera of the *Densovirus* subfamily, characterization of their expression strategies are still lacking. The major project of this thesis was the identification of the expression strategies of the members in *Iteradensovirus* genus, such as CeDNV and BmDNV-1, for which the complete genome organization were reported from our lab. Based on the discovery of these three new iteradensoviruses, we extended our scope to establish the transcription maps of these five iteradensoviruses.

A longstanding problem in densovirus research has been the lack of suitable cell lines to support viral genome replication. In our study of iteradensovirus transcription, we obtained the *Papilio polyxense* larvae for production of mRNA of PpDV. However, for the other four viruses, we lacked either host larvae or suitable cell lines to produce viral mRNA. Therefore, we attempted to transfet the infectious clones of these viruses into different insect cell lines, such as LD652, Sf9, C6/36 and Bmn, to determine the efficacy of mRNA production. RT-PCR results revealed that the LD652 cell line was the most efficient cell line for mRNA production from these four viruses. El-Far *et al.* (2003) reported LD652 showed high permissivity for both infection by the MIDNV purified virus and transfection by its infectious clone. However, the LD652 is not permissive to iteradensoviruses infection, the virions could not be rescued from the transfected cells even while the total RNA was enough for RT-PCR detection. Therefore, the LD652 cell line is suitable for viral transcription but not for virion production. Methods like MIMIC-PCR might be used to assay DNA replication and quantify genome copy numbers to establish the block point in the viral life cycle.

To investigate the expression strategy of iteradensovirus, we used Northern Blot to determine the transcript sizes of NS and VP genes. Both NS and VP probes revealed only single transcripts by Northern hybridization, which indicated that NS1 and NS2 may share the same polyadenylation motif at their 3' end. 3' RACE confirmed this result. The expression strategy of all five iteradensovirus were very similar to each other.

All the NS2 transcripts of iteradensovirus initiated from a highly conserved CAGT motif 5~6 nts downstream of the NS1 ATG. However, identifying the initiation site of NS1 transcripts was problematic due to the >40-fold excess of the NS2 transcript. Meanwhile, the potential initiation site of NS1 transcripts was quite close to the NS2 transcriptional start site and it was challenging to obtain the NS1 transcriptional start site by either conventional RACE method or RNase Protection Assay. Therefore, we designed a set of primers to obtain NS1-specific amplicons. Surprisingly, sequencing results showed the 5'-ends of NS1 transcripts did not start at the potential CATT/CATG/CATA motif 8 nts upstream of the NS1_{ATG}, but started from 2-4 nts upstream of the NS1_{ATG}. These imperfect transcription initiation sites of NS1 may also explain their much lower occurrence compared to NS2 transcripts. The NS1 and NS2 transcripts were not spliced and started 7-10 nts apart at the both side of NS1_{ATG}, which was similar to the NS transcription strategy of AalDNV (Pham *et al.*, 2013a). A TATA/TATA-box like motif can be found around 25 nts upstream of the NS1 initiator (Inr), but there was no obvious TATA motif upstream of the NS2 Inr core promoter sequence CAGT. It was reported that almost half of studied Drosophila promoters can be classified as TATA-less promoters (Arkhipova, 1995) and a conserved downstream core promoter element (DPE) has been recognized within TATA-less promoters in Drosophila, which is required for sequence-specific binding of the transcription factor TFIID. The DPE, with the consensus A/G/T-C/G-A/T-C/T-A/C/G-C/T acts in conjunction with the initiator sequence, Inr (CAGT), usually 28-32 nts downstream of the Inr, to provide a binding site for TFIID in the absence of a TATA box. Here, for the iteradensovirus, a DPE-like motif can be found downstream of NS2 Inr. Therefore, the transcription of iteradensovirus NS genes may be controlled by two core promoter

elements, the TATA-box or the DPE. To confirm this hypothesis, we expressed two NS proteins with an insect cell expression vector, replacing the original promoter with the potential iteradensovirus NS promoter elements. Serial mutations to remove the Inr or TATA motif were also created. The results revealed that the TATA-box is not an essential element of the NS2 promoter, the expression of the two NS proteins was controlled by two overlapping promoters that are responsible for different transcript start sites, similar to the brevidensovirus AaIDNV (Pham *et al.*, 2013a). Additional experiments are still required, such as the Luciferase promoter activity assay to characterize the detailed function of the NS promoters. At 3'-ends of iteradensovirus, NS1 and NS2 transcripts shared the same poly(A) motif which downstream of the TATA box of VP transcripts. This feature is similar as the ambidensoviruses and there is no exception among all the five iteradensoviruses. Although BmDNV contains an intergenic repeat between NS1 and VP transcripts, the poly(A) motif was still located downstream of VP TATA-box within the second repeated sequence. The strategy of iteradensoviruses NS and VP transcripts have individual poly(A) motif was different from the monosense brevidensovirus, the AaIDNV NS and VP transcripts share the same poly(A) motif at the 3'-ends of the virus genome. All the NS transcripts end around 10 nts downstream of the Inr of VP transcripts indicated that the 3'-ends of NS transcripts overlapped with the promoter elements of VP transcripts in monosense iteradensoviruses.

The iteradensoviruses do not have a complex expression strategy, the two NS proteins are transcribed from separate transcripts and the structural proteins are coded by unspliced transcripts that contain a single ORF. No splicing occurs for both NS and VP mRNAs and the different VP proteins are generated by a leaky scanning mechanism. Similar to the GmDNV (Tijssen *et al.*, 2003) and MIDNV (Fédière *et al.*, 2004) transcription strategies, the iteradensovirus VP transcripts also have a very short 5' untranslated region (10-15 nts). The short leader sequence of VP1 increases the probability that a fraction of the 40S ribosomal subunits will bypass its initiation codon to start at the next AUG initiation codon. TATA-box motifs are present upstream of the VP transcripts for all five iteradensoviruses. The poly(A) motif at the 3'-end overlaps

with the stop codons of the VP proteins which is unique among the other reported densoviruses' transcription maps. As for most vertebrate parvoviruses, iteradenosoviruses also contain a phospholipase A2 (PLA2) domain at the N-terminus of their VP1 proteins.

4 The unique transcription strategy of AdDNV

AdDNV has previously been classified into the *Pefudensovirus* genus of the *Densovirinae* subfamily according to its split VP ORFs which were similar to the *Periplaneta fuliginosa* densovirus (PfDNV) (Guo *et al.*, 2000). The newly established taxonomy of the *Densovirinae* subfamily has reclassified the PfDNV and AdDNV into the *Ambidensovirus* genus but within the separate *Blattodean ambidensovirus 1* and *Orthopteran ambisendovirus 1* species, respectively. We also determined a detailed transcriptional map of AdDNV, which is unique among the densoviruses.

The expression strategy of AdDNV nonstructural proteins (NS) was identical to the classical ambidensoviruses, such as GmDNV (Tijssen *et al.*, 2003) and MIDNV (Fédière *et al.*, 2004). Three NS proteins were expressed from either spliced or unspliced transcripts. However, the strategy of VP protein expression was completely unique among the densoviruses or vertebrate parvoviruses. The AdDNV VP gene consists of split ORFs and two sets of introns were found between them. The transcription of the AdDNV VP genes were not only dependent upon splicing, but also on leaky scanning strategies to express the different VP proteins. In BgDNV, the VP gene also consisted of two split ORFs, and the VP proteins were generated by alternative splicing. A strategy that is similar to the strategy used by AdDNV (Kapelinskaya *et al.*, 2011). The transcriptional start sites of the NS transcripts were localized at 23 nt and 573 nt downstream of the 5'-ITR, and the VP transcript was localized 23 nt upstream of the 3'-ITR, suggesting that the promoter element of both NS and VP genes is located within the ITRs, similar to other ambidensoviruses (e.g. GmDNV and MIDNV). In contrast, the promoter element of the NS gene for the

monosense iteradensovirus may only be partially located within the 5'-ITR due to the long distance between the ITRs and NS transcription start sites.

For most of the densoviruses, VP proteins were generated from a single ORF by a leaky scanning mechanism, their different VP proteins were resulted from different initiation sites for translation of the capsid gene and from posttranslational modification of their N termini. Therefore, each of the subunits that form the capsid usually contain the same C-terminal amino acid sequence but differ in their N-terminal extensions. In contrast, AdDNV VP2 was not completely contained within the corresponding VP1 as the alternative splicing to generate the different VPs resulted in split ORFs.

5 Comparison of AdDNV and sAAV capsid structures

To date, the capsid structures of several autonomous vertebrate parvovirus and densoviruses, as well as different serotypes of the adeno-associated virus (AAV) have been reported (Drouin *et al.*, 2013). *Parvovirinae* generally have three types of proteins (VP1, VP2, and VP3) in their capsids (Tijssen, 1999), whereas *Densovirinae* generally have four types of proteins (VP1 to VP4) in their capsids (Bergoin *et al.*, 2000). Due to the weak environment of VP1_{AUG}, the VP1 is usually generated as the minor protein of the four VPs of densoviruses (Tijssen *et al.*, 2003). For the members of *Parvovirinae*, the capsids are typically assembled from a few (5-10) copies of VP1 and 50 to 55 copies of various N-terminally truncated VP polypeptides (Cotmore *et al.*, 2014b). For instance, in AAV three VP proteins interact in a 1:1:10 ratio and the particles contain ~5 copies of VP1, ~5 copies of VP2, and ~50 copies of the shortest form, VP3; whereas MVM virions assemble from ~10 copies of VP1 and ~50 copies of VP2. In contrast, the four structural proteins of densovirus AdDNV VP1 (88.1 kDa), VP2 (65.3 kDa), VP3 (50.8 kDa), and VP4 (46.9 kDa) are in an approximate 1:11:18:30 proportion, indicating that there is only one copy of VP1 for capsid assembly.

In all parvoviruses, the BIDG sheet is 5-stranded with an N-terminal edge strand βA. The βA of AdDNV is associated with β-strand B as an N-terminal extension of the 2-fold related BIDG sheet. This structure is similar to other reported densoviruses which is an example of “domain swapping”. In contrast, the A β-strand of sAAV is folded back to run antiparallel to β-strand B in the BIDG sheet which is identical to all known vertebrate parvovirus structures (Chapman *et al.*, 2006). In AdDNV, the surface loops connecting the strands of the core jelly roll, the DE loop usually form the channel along the 5-fold icosahedral axes and include several hydrophobic amino acids that interact with the glycine-rich region. For sAAV, as well as the other serotype of AAVs, the loop regions inserted between β-strands consist of the distinctive HI loop between β-strands H and I, the DE loop between β-strands D and E, and nine variable regions (VRs), which form the top of the loops (Figure 6.2 B). Compared to other available AAV structures (Drouin *et al.*, 2013), VR-II is altered at the top of the 5-fold channel in sAAV. Meanwhile, sAAV showed the most significant difference at the 2-fold region due to an insertion and conformational change in the previously defined variable region (VR) IX which forms the wall of the depression (Figure 6.2 A). The VR-IV, VR-VIII, and VR-IX differences map to AAV regions reported to be determinants of receptor attachment, transduction determinants, and antigenic regions (Chapman *et al.*, 2006), and identify sAAV as a potential platform for developing AAV vectors which can be altered for specific cell targeting and host antibody immune evasion.

Most parvoviruses, including densoviruses, assemble *in vivo* both as full infectious particles and as empty particles. However, for AdDNV and presumably also for other densoviruses, the small fraction of particles that are empty in a virus preparation consist of only VP4 and are missing the glycine-rich sequence, whereas

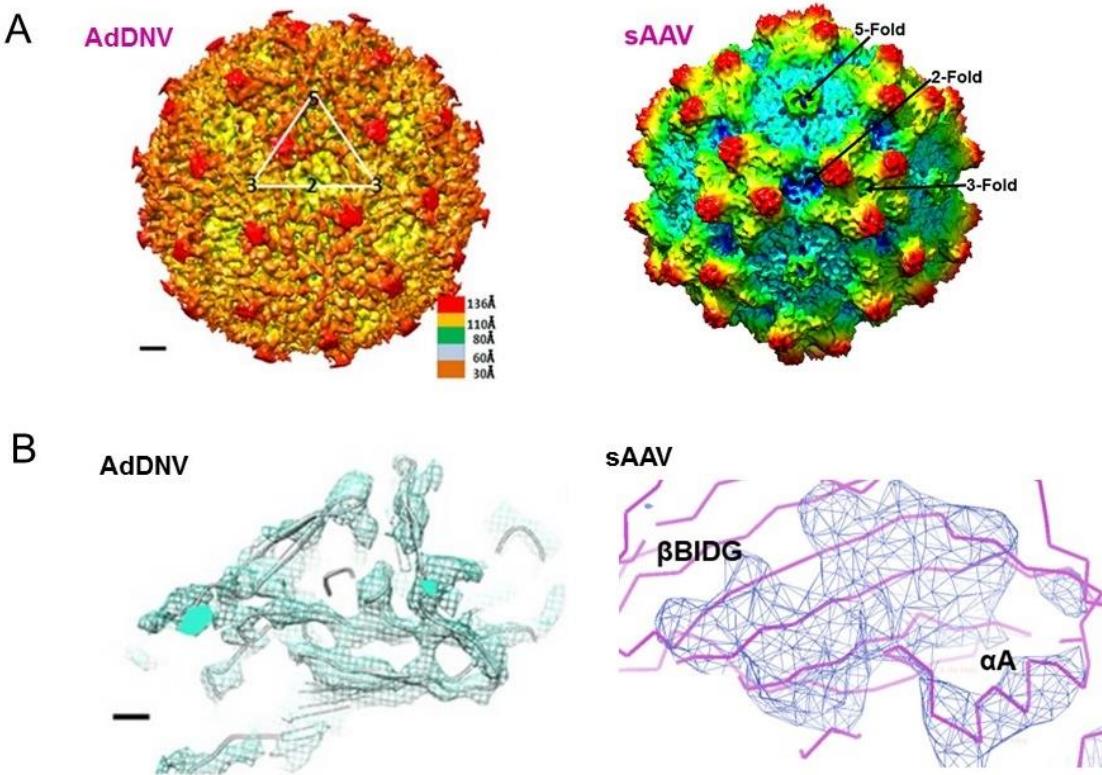


Figure 6.2 Structure comparison of AdDNV and sAAV.

A) Left: CryoEM reconstruction of emptied AdDNV particles. Surface features with a triangle showing the limits of one icosahedral asymmetric unit are shown. The scale bars represent 2 nm. Right: sAAV capsid structure. A conformational difference at VR-IX results in a wider 2-fold depression for sAAV. sAAV has less pronounced and rounder 3-fold protrusions due to change in VR-IV. B) Left: the BIDG β -sheet of AdDNV, the scale bars represent 3 Å. Right: the fit of sAAV α A into ordered tubular and β BIDG into sheet density (blue, threshold of $\sim 3\sigma$). (Meng et al., 2013) (Chipman et al., unpublished data)

the dominant infectious virus particles contain all four types of subunits (VP1 to VP4). Here, cryoEM demonstrated that heating of AdDNV for a defined length of time, increased the number of emptied particles as well as the PLA2 activity up till 65°C. When the temperature increased above 65°C, PLA2 activity reduced. Meanwhile, the VP1 N termini of the heat-treated emptied particles were sensitive to trypsin digestion, whereas the capsids remained intact as determined with cryoEM. This showed that the heat treatment causes externalization of the N-termini while leaving the capsid intact, with the PLA2 domain remaining a part of the particle. This could explain why all parvoviruses harbor the N-terminal part of VP1 within the virus particle until they are

ready to breach the endosomal membrane. This step is important for successful infection and it is believed to aid in viral escape from the endosome.

In general, we reported the capsid structures of AdDNV and sAAV and that externalization of the VP1 N-terminal region and subsequently genome release of AdDNV DNA by increasing temperature. These observations are significant to understand the evolutionary relationship between densoviruses and different AAV serotypes. Characterization of virus capsid 3D structures, enabled the identification of regions critical for particular tropisms and antigenic phenotypes.

6 The cricket densoviruses and other novel single-stranded DNA viruses of insects

Raising house crickets (*A. domesticus*) developed into a large-scale industry in Europe, Japan and North America since the 1940s (Weissman *et al.*, 2012). As the food source of reptiles such as lizards and snakes, and probably man, the cricket rearing business have been estimated to about \$650 million per year. In the pharmaceutical industry, crickets, instead of the American cockroach, are being used widely for physiological and toxicological studies. Crickets are very easy to rear in small or big scales. However, due to the density of populations, infectious diseases can spread quickly in these facilities. Besides the novel densoviruses AdMADV and segment AdSDV from crickets reported in chapter V, there are several cricket DNA viruses that have been reported during the past few years. This includes the cricket densovirus AdDNV (Meynardier *et al.*, 1977), and a novel vovolivirus AdVVV (Pham *et al.*, 2013e) we recently discovered from infected crickets samples from cricket facilities in the USA and Canada. Over the past two years, our lab has received many infected cricket samples for virus detection/identification and most of the samples have been positive for cricket viruses. These crickets were reported to have been shown similar visible symptoms of viral infection, such as anorexia, sluggishness, and slow death. As an increasingly important insect rearing industry in the USA and some European

countries, it is important to understand the diverse pathogens of crickets. Therefore, the detection and characterization of these novel viruses is increasingly valuable.

As a side project of this thesis, in appendix II, we also reported the genome sequences of three, novel single-stranded DNA viruses. The HaDNV and PiDNV for which we reported the complete genome sequences here were discovered long ago but the complete sequences were lacking. We also reported a new densovirus isolated from crickets and a circo-like single-stranded DNA virus from shrimp. This virus named PmCV-1 was the first circovirus found in shrimp. PmCV-1 is related to cycloviruses that have also been discovered in dragonflies, chickens, camels, cows, sheep, goats, and human stool samples. The presence of cycloviruses in a wide host range suggests that possible stable infection or latent infection might have been established in some of these hosts. Many viruses from shrimp have been reported before, including a large number of shrimp densoviruses which have been reclassified into two separate genera of *Densovirinae* subfamily (Cotmore *et al.*, 2014a). PstDNV is the major ssDNA virus of shrimp that causes significant disease in farmed and wild shrimp. During the last five decades, shrimp farming moved from small-scale enterprises in Southeast Asia into large-scale industry not only in Asia but also in South America. Therefore, the biological and molecular characteristics of different shrimp viruses will be important to limit their economic impact.

7 Horizontal transmission of cricket densovirus

AdDNV was the only densovirus reported to infect cricket. This virus was first isolated from diseased house cricket from a commercial facility in Switzerland (Meynardier *et al.*, 1977). Since then, it was found to cause serious epidemics in Germany, France, Netherlands, and England in the early 2000s. However, during the past two years, the detection of the AdDNV-negative cricket samples from different cricket farms in the USA and Canada showed another Brevidensovirus-like cricket virus might be responsible for the recent epidemic. Surprisingly, the genome of

this new cricket virus consisted of two similar-sized, segmented single-stranded DNA. These two segments contained the coding sequences for non-structural protein NS and structural protein VP separately. Both the NS- and VP- segment contained the same telomere sequences but were distinct at their left and right ends, a property shared with typical mosquito brevidensovirus. The amino acid sequences analysis showed that the NS-segment consist of two large ORFs coding for NS1 and NS2 proteins with high a.a. identity and gene organization with mosquito brevidensovirus. However, in the VP-segment, only the large ORF at N-terminal has high identity with mosquito brevidensovirus, whereas the C-terminus ORF has a high identity with the AdDNV VP protein. The highly conserved motifs of most densoviruses and vertebrate parvoviruses such as the RCR, ATPase and PLA2 motifs were also included in the NS-segment and VP-segment. There is no PLA2 motif in the typial mosquito brevidensoviruses. For the other densoviruses and vertebrate parvoriuses, the PLA2 motifs locate at the N-terminus of their VP proteins. The most interestingly feature of this new segment virus is the adaption of a PLA2 motif at the C-terminus of VP-segment. In chapter V, we also provided a hypothesis on the origin of this new segment brevidensovirus-like cricket virus. The virus was tentatively identified as *Acheta domesticus* segment densovirus (AdSDV). This is the first bipartite genome parvovirus that arose from mosquitoes and through recombination with AdDNV obtained a phospholipase A2 activity that enabled it to infect crickets. Meanwhile, the discovery of a novel cricket aphrodisiac iridescent virus which might contain a recombination with the AdDNV VP gene could be another example of horizontal transmission of the VP gene of cricket densovirus.

Pathogens offer many of the most fascinating and well-studied examples of evolution because of their speed of adaptation allows observation of evolutionary change within host lifetimes. The rapid evolution observed in pathogenic viruses, bacteria, and fungi affect human, animal, and plant health globally. For example, influenza viruses regularly recombine genes affecting host range, infection pathways, and virulence prior to the emergence of deadly outbreaks (Nelson *et al.*, 2007). Since the detection of cricket samples from different farms located both in Canada and the

USA, two different strains of the new segment cricket densovirus we reported here caused another epidemic within North America in the past two years. Although the mechanism of viral recombination between viruses of two different genera or families is still unclear, additional evidence will likely be discovered as more viral genomic data are generated and analyzed. A virus possessing a mechanism that facilitates such selective exchanges of host range determinants would have a major evolutionary advantage over competitors that lacked this capacity. Further analysis of the nature and consequences of the plasticity of the viral genomes could provide some important insights on how the evolutionary processes operate on cricket densoviruses.

CONCLUSIONS AND PERSPECTIVES

Densoviruses have a significant potential for biological control due to their highly pathogenicity in their natural hosts. Iteradensovirus CeDNV has been used to control the oil palm pest of the Ivory Coast. However, adding a virus to the biological control arsenal is a multi-step process which includes original virus isolation, molecular characterization an understanding of viral pathogenicity, and its safety towards non-target organisms. Therefore, the determination of densovirus' molecular characteristics including the genome organization, transcriptional strategy, and capsid structure is very important.

The transcription strategy and promoter structure of vertebrate parvoviruses has been more extensively studied than that of densoviruses. Although the transcript maps of some members in the *Densovirinae* subfamily have been fully investigated, none of the transcription strategies of members of *Iteradensovirus* genus has been reported previously. Here we first reported the cloning and sequencing of three new iteradensoviruses, extending the genus into seven members. And, for the first time, the expression strategies employed by five iteradensoviruses have been determined. There is no splicing for either NS or VP protein expression. Two NS proteins are generated by individual transcripts and controlled by overlapping promoters, similarly to the brevidensovirus AalDNV. All the VPs are generated from a single transcript by leaky scanning mechanisms. The expression strategy of the cricket ambisense densovirus AdDNV differed from the monosense iteradensoviruses, and its VP were generated by a unique alternative splicing mechanism combining two split ORFs as well as leaky scanning, a mechanism unique among all parvoviruses.

Among AdDNV molecular characteristics, we also described the crystal structure of mature virions at 3.5-Å resolution by X-ray crystallography and the cryo-electron microscopic (cryoEM) structure of the emptied virus at 5.5-Å resolution. We also reported the externalization of the VP1 N-terminal region and subsequent genome

release at increased temperatures. Furthermore, we compared the capsid structure of a snake parvovirus Serpine adeno-associated virus (sAAV). The sAAV contained all the capsid features of AAVs including a depression at the twofold axes, three protrusions surrounding the threefold axes, and a channel with surrounding canyon region at the fivefold axes of symmetry, but sAAV had significant differences within the variable regions.

An increasing number of novel densoviruses have been reported during the past few years. Methods such as the Sequence-independent single-primer amplification (SISPA) and deep sequencing have been developed for virus discovery. A novel cricket densovirus, AdMADV, and a novel shrimp circo-like ssDNA virus, PmCV-1, were discovered and sequenced. Most interestingly, a novel cricket densovirus (AdSDV), as the first bipartite genome parvovirus, has been discovered from the AdDNV-negative cricket samples. If primers to the N-terminus of AdDNV would have been included, then these samples could possibly have been AdDNV-positive since the PLA2 domain is well conserved. The virus arose from mosquitoes and through recombination with AdDNV obtained a PLA2 activity that enabled it to infect crickets. Such findings would highlight the intriguing possibility that novel virus groups can emerge via recombination between highly disparate virus types. In any case, the discovery of the AdSDV extends the modular theory of virus evolution to encompass a much broader range of possibilities than previously thought. Rearing crickets and shrimp are both very important industries, and hence identification of pathogens from infected crickets and shrimp thus provides more information for diagnosis, prevention, and control of emerging virus diseases.

Together, our results showed the densoviruses belonging to different genera of *Densovirinae* subfamily contained specific molecular characteristics to each genus. Their capsid structures include several common features as well as distinct differences with the vertebrate parvoviruses. Phylogeny analysis also demonstrated that vertebrate and invertebrate parvoviruses have evolved independently. Considering the

pathogenicity to their host, the densovirus should be significant candidates for biological control.

Single-stranded DNA viruses infect diverse prokaryotic and eukaryotic hosts, from plants, vertebrates and invertebrates and include important pathogens. Their small genomes usually replicate by the rolling-circle-like mechanism initiated by a distinct virus-encoded endonuclease. We investigated the provenance of the genes of AdSDV and uncovered an unexpected adaptation of a mosquito virus to crickets. Our analysis strongly suggests that the integration of a ambidensovirus PLA2 domain in the mosquito virus was critical for the adaptation to crickets. The unusual evolutionary history of this virus emphasizes the key role of horizontal gene transfer, sometimes between viruses with completely different genomes, as exemplified by the iridescent virus, but occupying the same niche, in the emergence of new viral types. A similar situation of a turbulent evolutionary history seems to exist for the *Bidnaviridae* (Krupovic *et al.*, 2014).

REFERENCES

- Adamo SA, Kovalko I, Easy RH & Stoltz D (2014) A viral aphrodisiac in the cricket *Gryllus texensis*. *The Journal of experimental biology* 217(11):1970-1976.
- Afanasiev BN, Galyov EE, Buchatsky LP & Kozlov YV (1991) Nucleotide sequence and genornic organization of *Aedes densonucleosis virus*. *Virology* 185(1):323-336.
- Agbandje M & Chapman MS (2006) *Correlating structure with function in the viral capsid*. Edward Arnold, New York, New York,
- Agbandje M, Kajigaya S, McKenna R, Young NS & Rossmann MG (1994) The Structure of Human Parvovirus B19 at 8 Å; Resolution. *Virology* 203(1):106-115.
- Agbandje M, Llamas-Saiz AL, Wang F, Tattersall P & Rossmann MG (1998) Functional implications of the structure of the murine parvovirus, minute virus of mice. *Structure* 6(11):1369-1381.
- Agbandje M, McKenna R, Rossmann MG, Strassheim ML & Parrish CR (1993) Structure determination of feline panleukopenia virus empty particles. *Proteins: Structure, Function, and Bioinformatics* 16(2):155-171.
- Ahne W & Scheinert P (1989) Reptilian Viruses: Isolation of Parvovirus - like Particles from Corn Snake *Elapha guttata* (Colubridae). *Journal of Veterinary Medicine* 36(1 - 10):409-412.
- Alford RA & Hammond A (1982) Plusiinae (Lepidoptera: Noctuidae) populations in Louisiana soybean ecosystems as determined with looplure-baited traps. *Journal of Economic Entomology* 75(4):647-650.
- Allander T, Emerson SU, Engle RE, Purcell RH & Bukh J (2001) A virus discovery method incorporating DNase treatment and its application to the identification of two bovine parvovirus species. *Proceedings of the National Academy of Sciences* 98(20):11609-11614.
- Anderson J & Brower L (1996) Freeze - protection of overwintering monarch butterflies in Mexico: critical role of the forest as a blanket and an umbrella. *Ecological Entomology* 21(2):107-116.
- Anonyme (2000) Virus taxonomy: classification and nomenclature of viruses. Seventh report of the International Committee on Taxonomy of Viruses. Van Regenmortel MH, Fauquet C, Bishop D, Carstens E, Estes M, Lemon S, Maniloff J, Mayo M, Mcgeoch D & Pringle C (Édit.) Academic Press.
- Arkhipova IR (1995) Promoter elements in *Drosophila melanogaster* revealed by sequence analysis. *Genetics* 139(3):1359-1369.

- Arnott HJ, Smith KM & Fullilove SL (1968) Ultrastructure of a cytoplasmic polyhedrosis virus affecting the monarch butterfly, *Danaus plexippus*: I. Development of virus and normal polyhedra in the larva. *Journal of Ultrastructure Research* 24(5):479-507.
- Asokan A, Conway JC, Phillips JL, Li C, Hegge J, Sinnott R, Yadav S, DiPrimio N, Nam H-J & Agbandje-McKenna M (2010) Reengineering a receptor footprint of adeno-associated virus enables selective and systemic gene transfer to muscle. *Nature Biotechnology* 28(1):79-82.
- Asokan A, Schaffer DV & Samulski RJ (2012) The AAV vector toolkit: poised at the clinical crossroads. *Molecular Therapy* 20(4):699-708.
- Astell CR, Mol CD & Anderson WF (1987) Structural and functional homology of parvovirus and papovavirus polypeptides. *Journal of General Virology* 68(Part 3):885-893.
- Bahar MW, Graham SC, Stuart DI & Grimes JM (2011) Insights into the evolution of a complex virus from the crystal structure of vaccinia virus D13. *Structure* 19(7):1011-1020.
- Baker ML, Zhang J, Ludtke SJ & Chiu W (2010) Cryo-EM of macromolecular assemblies at near-atomic resolution. *Nature Protocols* 5(10):1697-1708.
- Baquerizo-Audiot E, Abd-Alla A, Jousset F-X, Cousserans F, Tijssen P & Bergoin M (2009) Structure and expression strategy of the genome of *Culex pipiens* densovirus, a mosquito densovirus with an ambisense organization. *Journal of Virology* 83(13):6863-6873.
- Bartel RA, Oberhauser KS, De Roode JC & Altizer SM (2011) Monarch butterfly migration and parasite transmission in eastern North America. *Ecology* 92(2):342-351.
- Belloncik S (1989) Potential use of densonucleosis viruses as biological control agents of insect pests. *Handbook of Parvoviruses*, Tijssen P (Édit.) CRC press, Boca Raton Vol 2.
- Benson SD, Bamford JK, Bamford DH & Burnett RM (1999) Viral evolution revealed by bacteriophage PRD1 and human adenovirus coat protein structures. *Cell* 98(6):825-833.
- Berenbaum M & Feeny P (1981) Toxicity of angular furanocoumarins to swallowtail butterflies: escalation in a coevolutionary arms race? *Science* 212(4497):927-929.
- Bergeron J, Menezes J & Tijssen P (1993) Genomic organization and mapping of transcription and translation products of the NADL-2 strain of porcine parvovirus. *Virology* 197(1):86-98.
- Bergoin M & Tijssen P (2000) Molecular biology of Densovirinae. *Contrib. Microbiol.* 4:12-32.
- Bergoin M & Tijssen P (2008) Parvoviruses of arthropods. *Encyclopedia of Virology*. , Elsevier, Oxford, United Kingdom. p 76-85.

- Bergoin M & Tijssen P (2010) Densoviruses: a highly diverse group of arthropod parvoviruses. *Insect Virology*, Caister Academic Press, Great Britain. p 59-82.
- Bergvinson D & García-Lara S (2004) Genetic approaches to reducing losses of stored grain to insects and diseases. *Current Opinion in Plant Biology* 7(4):480-485.
- Berns K (1990) Parvovirus replication. *Microbiological Reviews* 54(3):316-329.
- Berns K & Adler S (1972) Separation of two types of adeno-associated virus particles containing complementary polynucleotide chains. *Journal of Virology* 9(2):394.
- Berns K, Bergoin M, Bloom M, Lederman M, Muzyczka N, Siegl G, Tal J & Tattersall P (2000) *Family parvoviridae*. 311-323 p
- Berns K & Linden RM (1995) The cryptic life style of adenoassociated virus. *Bioessays* 17(3):237-245.
- Berns K & Parrish CR (2006) *Parvoviridae. Fields Virology*, Knipe D & Howley P (Édit.) Philadelphia, PA. p 2437-2478.
- Berns K & Parrish CR (2013) *Parvoviridae. Fields Virology*, Knipe D & Howley P (Édit.) Lippincott Williams & Wilkins, Philadelphia, 6th Ed.
- Boisvert M, Fernandes S & Tijssen P (2010) Multiple pathways involved in porcine parvovirus cellular entry and trafficking toward the nucleus. *Journal of Virology* 84(15):7782-7792.
- Bonning BC & Hammock BD (1996) Development of recombinant baculoviruses for insect control. *Annual Review of Entomology* 41(1):191-210.
- Boublik Y, Jousset F-X & Bergoin M (1994) Complete Nucleotide Sequence and Genomic Organization of the *Aedes albopictus* Parvovirus (AaPV) Pathogenic for *Aedes aegypti* Larvae. *Virology* 200(2):752-763.
- Brower LP, Fink LS & Walford P (2006) Fueling the fall migration of the monarch butterfly. *Integrative and Comparative Biology* 46(6):1123-1142.
- Brunger AT, Adams PD, Clore GM, DeLano WL, Gros P, Grosse-Kunstleve RW, Jiang J-S, Kuszewski J, Nilges M & Pannu NS (1998) Crystallography & NMR system: a new software suite for macromolecular structure determination. *Acta Crystallographica Section D: Biological Crystallography* 54(5):905-921.
- Buchatsky L, Bogdanova E, Kuznetsova M, Lebedinets N, Konongo A, Chabanenko A & Podrezova L (1987) Field tests of the viral preparation viroden on preimago stages blood-sucking mosquitoes. *Med Parazitol* 4:69-72.
- Canaan S, Zádori Z, Ghomashchi F, Bollinger J, Sadilek M, Moreau ME, Tijssen P & Gelb MH (2004) Interfacial enzymology of parvovirus phospholipases A2. *Journal of Biological Chemistry* 279(15):14502-14508.

- Carlini CR & Grossi-de-Sá MF (2002) Plant toxic proteins with insecticidal properties. A review on their potentialities as bioinsecticides. *Toxicon* 40(11):1515-1539.
- Carlson J, Suchman E & Buchatsky L (2006) Densovirus for control and genetic manipulation of mosquitoes. *AVR* 68:361-392.
- Chao Y-C, Young III S, Kim K & Scott H (1985) A newly isolated denso-nucleosis virus from *Pseudoplusia includens* (Lepidoptera: Noctuidae). *Journal of Invertebrate Pathology* 46(1):70-82.
- Chapman MS & Agbandje-McKenna M (2006) Atomic structure of viral particles. Hodder Arnold, London. *Parvoviruses*:107-123p.
- Chapman MS & Rossmann MG (1993) Structure, sequence, and function correlations among parvoviruses. *Virology* 194(2):491-508.
- Chen, S., Cheng, L., Zhang, Q., Lin, W., Lu, X., Brannan, J., Zhou, Z. H. & Zhang, J. (2004). Genetic, biochemical, and structural characterization of a new densovirus isolated from a chronically infected *Aedes albopictus* C6/36 cell line. *Virology* 318, 123–133.
- Collaborative CP (1994) The CCP4 suite: programs for protein crystallography. *Acta Crystallographica* 50(Pt 5):760.
- Cossons N, Faust EA & Zannis-Hadjopoulos M (1996) DNA polymerase δ -dependent formation of a hairpin structure at the 5' terminal palindrome of the minute virus of mice genome. *Virology* 216(1):258-264.
- Cotmore S, Agbandje-McKenna M, Chiorini JA, Mukha DV, Pintel DJ, Qiu J, Soderlund-Venermo M, Tattersall P, Tijssen P & Gatherer D (2014a) The family *Parvoviridae*. *Archives of Virology* 159(5):1239-1247.
- Cotmore S, D'Abramo AM, Jr., Carbonell LF, Bratton J & P. T (1997) The NS2 polypeptide of parvovirus MVM is required for capsid assembly in murine cells. *Virology* 231(2):267-280.
- Cotmore S & Tattersall P (1989) A genome-linked copy of the NS-1 polypeptide is located on the outside of infectious parvovirus particles. *Journal of Virology* 63(9):3902-3911.
- Cotmore S & Tattersall P (2006a) *Structure and organization of the viral genome*. Oxford University Press, London, U.K. . 73-94 p
- Cotmore S & Tattersall P (2006b) A rolling-hairpin strategy: basic mechanisms of DNA replication in the parvoviruses. *Parvoviruses*, Kerr J, Cotmore S, Bloom M, Linden R & Parrish CR (Édit.) Hodder Arnold, London. p 171-188.
- Cotmore S & Tattersall P (2006c) Parvoviruses. *DNA Replication and Human Disease* Depamphilis M (Édit.) Cold Spring Harb. Lab. Press, New York. p 593-608.

- Cotmore S & Tattersall P (2013) Parvovirus diversity and DNA damage responses. *Cold Spring Harbor perspectives in biology* 5(2):a012989.
- Cotmore S & Tattersall PJ (2014b) Parvoviruses: Small Does Not Mean Simple. *Annual Review of Virology* 1(1):517-537.
- Courey AJ & Tjian R (1988) Analysis of Sp1 in vivo reveals multiple transcriptional domains, including a novel glutamine-rich activation motif. *Cell* 55(5):887-898.
- Croizier L, Jousset F-X, Veyrunes J-C, López-Ferber M, Bergoin M & Croizier G (2000) Protein requirements for assembly of virus-like particles of *Junonia coenia* densovirus in insect cells. *Journal of General Virology* 81(6):1605-1613.
- Crook NE, Clem R & Miller L (1993) An apoptosis-inhibiting baculovirus gene with a zinc finger-like motif. *Journal of Virology* 67(4):2168-2174.
- De Bach P (1964) *Biological control of insect pests and weeds*. Chapman & Hall, London.
- De Roode J, de CASTILLEJO F, LOPEZ C, Faits T & Alizon S (2011) Virulence evolution in response to anti - infection resistance: toxic food plants can select for virulent parasites of monarch butterflies. *Journal of Evolutionary Biology* 24(4):712-722.
- de Vega M, Lázaro JM, Mencía M, Blanco L & Salas M (2010) Improvement of φ29 DNA polymerase amplification performance by fusion of DNA binding motifs. *Proceedings of the National Academy of Sciences* 107(38):16506-16511.
- Delwart E & Li L (2012) Rapidly expanding genetic diversity and host range of the *Circoviridae* viral family and other Rep encoding small circular ssDNA genomes. *Virus Research* 164(1):114-121.
- Dhar AK, Kaizer KN & Lakshman DK (2010) Transcriptional analysis of *Penaeus stylirostris* densovirus genes. *Virology* 402(1):112-120.
- DiMattia MA, Nam H-J, Van Vliet K, Mitchell M, Bennett A, Gurda BL, McKenna R, Olson NH, Sinkovits RS & Potter M (2012) Structural insight into the unique properties of adeno-associated virus serotype 9. *Journal of Virology* 86(12):6947-6958.
- Drouin LM & Agbandje-McKenna M (2013) Adeno-associated virus structural biology as a tool in vector development. *Future Virology* 8(12):1183-1199.
- Dumas B, Jourdan M, Pascaud A-M & Bergoin M (1992) Complete nucleotide sequence of the cloned infectious genome of *Junonia coenia* densovirus reveals an organization unique among parvoviruses. *Virology* 191(1):202-222.
- Dunlap DS, Ng TFF, Rosario K, Barbosa JG, Greco AM, Breitbart M & Hewson I (2013) Molecular and microscopic evidence of viruses in marine copepods. *Proceedings of the National Academy of Sciences* 110(4):1375-1380.

- El-Far M, Szelei J, Yu Q, Fédière G, Bergoin M & Tijssen P (2012) Organization of the Ambisense Genome of the *Helicoverpa armigera* Densovirus. *Journal of Virology* 86(12):7024-7024.
- Emsley P & Cowtan K (2004) Coot: model-building tools for molecular graphics. *Acta Crystallographica Section D: Biological Crystallography* 60(12):2126-2132.
- Erles K, Šebeková P & Schlehofer JR (1999) Update on the prevalence of serum antibodies (IgG and IgM) to adeno - associated virus (AAV). *Journal of Medical Virology* 59(3):406-411.
- Farkas SL, Zádori Z, Benkő M, Essbauer S, Harrach B & Tijssen P (2004) A parvovirus isolated from royal python (*Python regius*) is a member of the genus *Dependovirus*. *Journal of General Virology* 85(3):555-561.
- Farr GA, Zhang L-g & Tattersall P (2005) Parvoviral virions deploy a capsid-tethered lipolytic enzyme to breach the endosomal membrane during cell entry. *Proceedings of the National Academy of Sciences of the United States of America* 102(47):17148-17153.
- Fédière G (1996) *Recherches sur des viroses de l'épidoptères ravageurs de cultures pérennes en Côte d'Ivoire et de cultures annuelles en Egypte*.
- Fédière G (1983) Recherches sur des viroses épidémiologiques de l'épidoptères Limacodidae ravageurs de palmiers.
- Fédière G (2000) Epidemiology and pathology of Densovirinae. *Parvoviruses*, (From molecular biology to pathology and therapeutic uses., Faisst S & Rommelaere J (Édit.) Karger, Basel, Switzerland. p 1-11.
- Fédière G, El-Far M, Li Y, Bergoin M & Tijssen P (2004) Expression strategy of densonucleosis virus from *Mythimna loreyi*. *Virology* 320(1):181-189.
- Fédière G, El-Sheikh M, Abol-Ela S, Salah M, Massri M & Veyrunes J (1995) Isolation of a new densonucleosis virus from *Mythimna loreiyi* Dup.(Lep. Noctuidae) in Egypt. *Bulletin of Faculty of Agriculture* 46(4):693-702.
- Fédière G, Li Y, Zadori Z, Szelei J & Tijssen P (2002) Genome Organization of *Caspalia extranea* Densovirus, a New *Iteravirus*. *Virology* 292(2):299-308.
- Fernandes S & Tijssen P (2009) Seamless cloning and domain swapping of synthetic and complex DNA. *AB* 385(1):171-173.
- Fest C & Schmidt K-J (1982) *The chemistry of organophosphorus pesticides*. Springer-Verlag, New York
- Firth C & Lipkin WI (2013) The genomics of emerging pathogens. *Annual Review of Genomics and Human Genetics* 14:281-300.

Flint SJ, Racaniello V, Enquist L & Skalka AM (2009) Genome replication strategies: DNA viruses. *Principles of Virology*, Ed. 3), Flint SJ, Racaniello V, Enquist L & Skalka AM (Édit.) ASM press, Washington, D.C.

Garriga D, Pickl-Herk A, Luque D, Wruss J, Castón JR, Blaas D & Verdaguer N (2012) Insights into minor group rhinovirus uncoating: the X-ray structure of the HRV2 empty capsid. *PLoS Pathogens* 8(1):e1002473.

Genty P & Mariau D (1975) Utilisation d'un germe entomopathogène dans la lutte contre *Sibine fusca* (*Limacodidae*). *Oleagineux* 30:349-354.

Girod A, Wobus CE, Zádori Z, Ried M, Leike K, Tijssen P, Kleinschmidt JA & Hallek M (2002) The VP1 capsid protein of adeno-associated virus type 2 is carrying a phospholipase A2 domain required for virus infectivity. *Journal of General Virology* 83(5):973-978.

Godiska R, Mead D, Dhodda V, Wu C, Hochstein R, Karsi A, Usdin K, Entezam A & Ravin N (2009) Linear plasmid vector for cloning of repetitive or unstable sequences in *Escherichia coli*. *Nucleic Acids Research*:gkp1181.

Goodwin R (1985) Growth of insect cells in serum-free media. . *Techniques in the life sciences*, Elsevier Scientific Publishers, Ireland Ltd. p 1-28.

Govindasamy L, DiMattia MA, Gurda BL, Halder S, McKenna R, Chiorini JA, Muzyczka N, Zolotukhin S & Agbandje-McKenna M (2013) Structural insights into adeno-associated virus serotype 5. *Journal of Virology* 87(20):11187-11199.

Govindasamy L, Padron E, McKenna R, Muzyczka N, Kaludov N, Chiorini JA & Agbandje-McKenna M (2006) Structurally mapping the diverse phenotype of adeno-associated virus serotype 4. *Journal of Virology* 80(23):11556-11570.

Gudenkauf BM, Eaglesham JB, Aragundi WM & Hewson I (2014) Discovery of urchin-associated densoviruses (family *Parvoviridae*) in coastal waters of the Big Island, Hawaii. *Journal of General Virology* 95(Pt 3):652-658.

Gunning RV, Dang HT, Kemp FC, Nicholson IC & Moores GD (2005) New resistance mechanism in *Helicoverpa armigera* threatens transgenic crops expressing *Bacillus thuringiensis* Cry1Ac toxin. *AEM* 71(5):2558-2563.

Guo H, Zhang J & Hu Y (2000) Complete sequence and organization of *Periplaneta fuliginosa* densovirus genome. *AV* 44(6):315-322.

Gurda BL, DiMattia MA, Miller EB, Bennett A, McKenna R, Weichert WS, Nelson CD, Chen W-j, Muzyczka N & Olson NH (2013) Capsid antibodies to different adeno-associated virus serotypes bind common regions. *Journal of Virology* 87(16):9111-9124.

Halder S, Nam HJ, Govindasamy L, Vogel M, Dinsart C, Salome N, McKenna R & Agbandje-McKenna M (2013) Structural characterization of H-1 parvovirus: comparison of infectious virions to empty capsids. *Journal of Virology* 87(9):5128-5140.

- Harrison S, Olson A, Schutt C, Winkler F & Bricogne G (1978) Tomato bushy stunt virus at 2.9 Å resolution.
- Harvell C, Kim K, Burkholder J, Colwell R, Epstein PR, Grimes D, Hofmann E, Lipp E, Osterhaus A & Overstreet RM (1999) Emerging marine diseases--climate links and anthropogenic factors. *Science* 285(5433):1505-1510.
- Hellman U, Wernstedt C, Gonez J & Heldin C-H (1995) Improvement of an "In-Gel" digestion procedure for the micropreparation of internal protein fragments for amino acid sequencing. *AB* 224(1):451-455.
- Herniou EA, Olszewski JA, Cory JS & O'Reilly DR (2003) The genome sequence and evolution of baculoviruses. *Annual Review of Entomology* 48(1):211-234.
- Higgins DG & Sharp PM (1988) CLUSTAL: a package for performing multiple sequence alignment on a microcomputer. *Gene* 73(1):237-244.
- Hogle J, Chow M & Filman D (1985) Three-dimensional structure of poliovirus at 2.9 Å resolution. *Science* 229(4720):1358-1365.
- Holton T & Graham M (1991) A simple and efficient method for direct cloning of PCR products using ddT-tailed vectors. *Nucleic Acids Research* 19(5):1156.
- Hörer M, Weger S, Butz K, Hoppe-Seyler F, Geisen C & Kleinschmidt JA (1995) Mutational analysis of adeno-associated virus Rep protein-mediated inhibition of heterologous and homologous promoters. *Journal of Virology* 69(9):5485-5496.
- Hsu J-Y, Juven-Gershon T, Marr MT, Wright KJ, Tjian R & Kadonaga JT (2008) TBP, Mot1, and NC2 establish a regulatory circuit that controls DPE-dependent versus TATA-dependent transcription. *Acta Crystallographica. Section D, Biological Crystallography Genes & Development* 22(17):2353-2358.
- Huang X & Madan A (1999) CAP3: A DNA sequence assembly program. *Genome Research* 9(9):868-877.
- Huotari J & Helenius A (2011) Endosome maturation. *The EMBO journal* 30(17):3481-3500.
- Huynh OT, Pham HT, Yu Q & Tijssen P (2012) *Pseudoplusia includens* densovirus genome organization and expression strategy. *Journal of Virology* 86(23):13127-13128.
- Inceoglu AB, Kamita SG, Hinton AC, Huang Q, Severson TF, Kang Kd & Hammock BD (2001) Recombinant baculoviruses for insect control. *Pest Management Science* 57(10):981-987.
- Jourdan M, Jousset F-X, Gervais M, Skory S, Bergoin M & Dumas B (1990) Cloning of the genome of a densovirus and rescue of infectious virions from recombinant plasmid in the insect host *Spodoptera littoralis*. *Virology* 179(1):403-409.

- Kadonaga JT (2002) The DPE, a core promoter element for transcription by RNA polymerase II. *Experimental and Molecular Medicine* 34(4):259-264.
- Kapelinskaya T, Martynova E, Korolev A, Schal C & Mukha D (2008) Transcription of the German cockroach densovirus BgDNV genome: alternative processing of viral RNAs. *Doklady Biochemistry and Biophysics*. Springer, p 176-180.
- Kapelinskaya T, Martynova EU, Schal C & Mukha DV (2011) Expression strategy of densonucleosis virus from the German cockroach, *Blattella germanica*. *Journal of Virology* 85(22):11855-11870.
- Kaufmann B, Bowman VD, Li Y, Szelei J, Waddell PJ, Tijssen P & Rossmann MG (2010) Structure of *Penaeus stylirostris* densovirus, a shrimp pathogen. *Journal of Virology* 84(21):11289-11296.
- Kaufmann B, El-Far M, Plevka P, Bowman VD, Li Y, Tijssen P & Rossmann MG (2011) Structure of *Bombyx mori* densovirus 1, a silkworm pathogen. *Journal of Virology* 85(10):4691-4697.
- Kaufmann B, Simpson AA & Rossmann MG (2004) The structure of human parvovirus B19. *Proceedings of the National Academy of Sciences of the United States of America* 101(32):11628-11633.
- Kawase S (1985) Pathology associated with densoviruses. *Viral Insecticifrd for Biological Control*, K. M & E. S (Édit.) Academic Press, New York. p 197-231.
- Kawase S, Garzon S, Su D & Tijssen P (1990) Insect parvovirus diseases. *Handbook of parvoviruses*, Tijssen P (Édit.) CRC Press, Vol VII. p 213-228.
- Kawase S & Seok K (1976) On the nucleic acid of a newly-isolated virus from the flacherie-diseased silkworm larvae. *Journal of Sericultural Science of Japan*.
- Kleywegt GJ & Read RJ (1997) Not your average density. *Structure* 5(12):1557-1569.
- Knight A & Brower LP (2009) The influence of eastern North American autumnal migrant monarch butterflies (*Danaus plexippus L.*) on continuously breeding resident monarch populations in southern Florida. *Journal of Chemical Ecology* 35(7):816-823.
- Kojima KK, Matsumoto T & Fujiwara H (2005) Eukaryotic translational coupling in UAAUG stop-start codons for the bicistronic RNA translation of the non-long terminal repeat retrotransposon SART1. *Molecular and Cellular Biology* 25(17):7675-7686.
- Kotchey NM, Adachi K, Zahid M, Inagaki K, Charan R, Parker RS & Nakai H (2011) A potential role of distinctively delayed blood clearance of recombinant adeno-associated virus serotype 9 in robust cardiac transduction. *Molecular Therapy* 19(6):1079-1089.

- Kronenberg S, Kleinschmidt JA & B ötcher B (2001) Electron cryo-microscopy and image reconstruction of adeno-associated virus type 2 empty capsids. *EMBO reports* 2(11):997-1002.
- Krupovic M & Koonin EV (2014) Evolution of eukaryotic single-stranded DNA viruses of the *Bidnaviridae* family from genes of four other groups of widely different viruses. *Scientific reports* 4.
- Kurstak E (1982) *Microbial and viral pesticides*. Marcel Dekker, New York, Bessel
- Laemmli UK (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* 227(5259):680-685.
- Lange M & Jehle JA (2003) The genome of the *Cryptophlebia leucotreta* granulovirus. *Virology* 317(2):220-236.
- Lavine M & Beckage N (1995) Polydnaviruses: potent mediators of host insect immune dysfunction. *Parasitology Today* 11(10):368-378.
- Legendre D & Rommelaere J (1992) Terminal regions of the NS-1 protein of the parvovirus minute virus of mice are involved in cytotoxicity and promoter trans inhibition. *Journal of Virology* 66(10):5705-5713.
- Legendre D & Rommelaere J (1994) Targeting of promoters for trans activation by a carboxy-terminal domain of the NS-1 protein of the parvovirus minute virus of mice. *Journal of Virology* 68(12):7974-7985.
- Lech TF, Xie Q & Chapman MS (2010) The structure of adeno-associated virus serotype 3B (AAV-3B): insights into receptor binding and immune evasion. *Virology* 403(1):26-36.
- Li L, Cotmore SF & Tattersall P (2013) Parvoviral left-end hairpin ears are essential during infection for establishing a functional intranuclear transcription template and for efficient progeny genome encapsidation. *Journal of Virology* 87(19):10501-10514.
- Li Y, Zadori Z, Bando H, Dubuc R, Fédi ère G, Szelei J & Tijssen P (2001) Genome organization of the densovirus from *Bombyx mori* (BmDNV-1) and enzyme activity of its capsid. *Journal of General Virology* 82(11):2821-2825.
- Liu K, Li Y, Jousset F-X, Zadori Z, Szelei J, Yu Q, Pham HT, L épine F, Bergoin M & Tijssen P (2011) The *Acheta domesticus* densovirus, isolated from the European house cricket, has evolved an expression strategy unique among parvoviruses. *Journal of Virology* 85(19):10069-10078.
- Llamas-Saiz A, Agbandje-McKenna M, Wikoff W, Bratton J, Tattersall P & Rossmann M (1997) Structure determination of minute virus of mice. *Acta Crystallographica Section D: Biological Crystallography* 53(1):93-102.

- Longworth J, Tinsley T, Barwise A & Walker I (1968) Purification of a Non-occluded Virus of *Galleria mellonella*. *Journal of General Virology* 3(2):167-174.
- Lorson C, Burger LR, Mouw M & Pintel DJ (1996) Efficient transactivation of the minute virus of mice P38 promoter requires upstream binding of NS1. *Journal of Virology* 70(2):834-842.
- Lou HJ, Brister JR, Li JJ, Chen W, Muzyczka N & Tan W (2004) Adeno-Associated Virus Rep78/Rep68 Promotes Localized Melting of the Rep Binding Element in the Absence of Adenosine Triphosphate. *ChemBioChem* 5(3):324-332.
- Ludtke SJ, Baldwin PR & Chiu W (1999) EMAN: semiautomated software for high-resolution single-particle reconstructions. *Journal of Structural Biology* 128(1):82-97.
- Lusby E, Fife K & Berns K (1980) Nucleotide sequence of the inverted terminal repetition in adeno-associated virus DNA. *Journal of Virology* 34(2):402-409.
- Macrae TC, Baur ME, Boethel DJ, Fitzpatrick BJ, Gao A-G, Gamundi JC, Harrison LA, Kabuye VT, Mcpherson RM & Miklos JA (2005) Laboratory and field evaluations of transgenic soybean exhibiting high-dose expression of a synthetic *Bacillus thuringiensis* cry1A gene for control of Lepidoptera. *Journal of Economic Entomology* 98(2):577-587.
- Malcolm SB, Cockrell BJ & Brower LP (1989) Cardenolide fingerprint of monarch butterflies reared on common milkweed, *Asclepias syriaca* L. *Journal of Chemical Ecology* 15(3):819-853.
- Mani B, Baltzer C, Valle N, Almendral JM, Kempf C & Ros C (2006) Low pH-dependent endosomal processing of the incoming parvovirus minute virus of mice virion leads to externalization of the VP1 N-terminal sequence (N-VP1), N-VP2 cleavage, and uncoating of the full-length genome. *Journal of Virology* 80(2):1015-1024.
- Matsudaira P (1987) Sequence from picomole quantities of proteins electroblotted onto polyvinylidene difluoride membranes. *Journal of Biological Chemistry* 262(21):10035-10038.
- McPherson RM & MacRae TC (2009) Evaluation of transgenic soybean exhibiting high expression of a synthetic *Bacillus thuringiensis* cry1A transgene for suppressing lepidopteran population densities and crop injury. *Journal of Economic Entomology* 102(4):1640-1648.
- Meng G, Zhang X, Plevka P, Yu Q, Tijssen P & Rossmann MG (2013) The structure and host entry of an invertebrate parvovirus. *Journal of Virology* 87(23):12523-12530.
- Meynadier G, Vago C, Plantevin G & Atger P (1964) Virose d'un type inhabituel chez le lépidoptère *Galleria mellonella* L. *Rev. Zool. Agric. Appl* 63:207-208.
- Meynardier G & al. e (1977) Virus de type densonucléose chez les orthoptères. *Ann. Soc. Entomol. Fr.* 13:487-493.

- Mingozzi F & High KA (2011) Therapeutic in vivo gene transfer for genetic disease using AAV: progress and challenges. *Nature Reviews Genetics* 12(5):341-355.
- Mukha D, Chumachenko A, Dykstra M, Kurtti T & Schal C (2006) Characterization of a new densovirus infecting the German cockroach, *Blattella germanica*. *Journal of General Virology* 87(6):1567-1575.
- Mukha D & Schal K (2003) [A densovirus of German cockroach *Blatella germanica*: detection, nucleotide sequence and genome organization]. *Molekuliarnaia Biologiiia* 37(4):607-618.
- Muzyczka N & Berns K (2001) Parvoviridae: the viruses and their replication. *Fields Virology* 2:2327-2359.
- Naccache SN, Greninger AL, Lee D, Coffey LL, Phan T, Rein-Weston A, Aronsohn A, Hackett J, Delwart EL & Chiu CY (2013) The perils of pathogen discovery: origin of a novel parvovirus-like hybrid genome traced to nucleic acid extraction spin columns. *Journal of Virology* 87(22):11966-11977.
- Nakagaki M & Kawase S (1980) Structural proteins of densonucleosis virus isolated from the silkworm, *Bombyx mori*, infected with the flacherie virus. *Journal of Invertebrate Pathology* 36(2):166-171.
- Nam H-J, Gurda BL, McKenna R, Potter M, Byrne B, Salganik M, Muzyczka N & Agbandje-McKenna M (2011) Structural studies of adeno-associated virus serotype 8 capsid transitions associated with endosomal trafficking. *Journal of Virology* 85(22):11791-11799.
- Nam H-J, Lane MD, Padron E, Gurda B, McKenna R, Kohlbrenner E, Aslanidi G, Byrne B, Muzyczka N & Zolotukhin S (2007) Structure of adeno-associated virus serotype 8, a gene therapy vector. *Journal of Virology* 81(22):12260-12271.
- Nandhagopal N, Simpson AA, Gurnon JR, Yan X, Baker TS, Graves MV, Van Etten JL & Rossmann MG (2002) The structure and evolution of the major capsid protein of a large, lipid-containing DNA virus. *Proceedings of the National Academy of Sciences* 99(23):14758-14763.
- Nelson MI & Holmes EC (2007) The evolution of epidemic influenza. *Nature reviews genetics* 8(3):196-205.
- Ng R, Govindasamy L, Gurda BL, McKenna R, Kozyreva OG, Samulski RJ, Parent KN, Baker TS & Agbandje-McKenna M (2010) Structural characterization of the dual glycan binding adeno-associated virus serotype 6. *Journal of Virology* 84(24):12945-12957.
- Nicholson GM (2007) Fighting the global pest problem: preface to the special Toxicon issue on insecticidal toxins and their potential for insect pest control. *Toxicon* 49(4):413-422.
- Niiler E (1999) GM corn poses little threat to monarch. *Nat. Biotechnol* 17(12):1.154.

- Novotny V, Basset Y, Miller SE, Weiblen GD, Bremer B, Cizek L & Drozd P (2002) Low host specificity of herbivorous insects in a tropical forest. *Nature* 416(6883):841-844.
- Nüesch J (2006) Regulation of non-structural protein functions by differential synthesis, modification and trafficking. *Parvoviruses*, Jonathan K, Susan C & E BM (Édit.). p 275-289.
- Oerke E-C & Dehne H-W (2004) Safeguarding production—losses in major crops and the role of crop protection. *Crop Protection* 23(4):275-285.
- Opie SR, Warrington Jr KH, Agbandje-McKenna M, Zolotukhin S & Muzychka N (2003) Identification of amino acid residues in the capsid proteins of adeno-associated virus type 2 that contribute to heparan sulfate proteoglycan binding. *Journal of Virology* 77(12):6995-7006.
- Otwinowski Z & Minor W (1997) Processing of X-ray diffraction data. *Methods Enzymol* 276:307-326.
- Paterson A, Robinson E, Suchman E, Afanasiev B & Carlson J (2005) Mosquito densonucleosis viruses cause dramatically different infection phenotypes in the C6/36 *Aedes albopictus* cell line. *Virology* 337(2):253-261.
- Pham HT, Bergoin M & Tijssen P (2013c) *Acheta domesticus* volvovirus, a novel single-stranded circular DNA virus of the house cricket. *Genome Announcements* 1(2):e00079-00013.
- Pham HT, Huynh OT, Jousset F-X, Bergoin M & Tijssen P (2013d) *Junonia coenia* densovirus (JcDNV) genome structure. *Genome Announcements* 1(4):e00591-00513.
- Pham HT, Iwao H, Bergoin M & Tijssen P (2013e) New volvovirus isolates from *Acheta domesticus* (Japan) and *Gryllus assimilis* (United States). *Genome Announcements* 1(3):e00328-00313.
- Pham HT, Iwao H, Szelei J, Li Y, Liu K, Bergoin M & Tijssen P (2013b) Comparative genomic analysis of *Acheta domesticus* densovirus isolates from different outbreaks in Europe, North America, and Japan. *Genome Announcements* 1(4):e00629-00613.
- Pham HT, Jousset F-X, Perreault J, Shike H, Szelei J, Bergoin M & Tijssen P (2013a) Expression Strategy of *Aedes albopictus* Densovirus. *Journal of Virology* 87(17):9928-9932.
- Pham HT, Yu Q, Bergoin M & Tijssen P (2013f) A Novel Ambisense Densovirus, *Acheta domesticus* Mini Ambidensovirus, from Crickets. *Genome Announcement* 1(6).
- Powell ML, Brown TD & Brierley I (2008a) Translational termination-re-initiation in viral systems. *Biochemical Society Transactions* 36(4):717-722.

- Powell ML, Napthine S, Jackson RJ, Brierley I & Brown TDK (2008b) Characterization of the termination–reinitiation strategy employed in the expression of influenza B virus BM2 protein. *RNA* 14(11):2394-2406.
- Qiu J, Yoto Y, Tullis G & Pintel D (2006) *Parvovirus RNA processing strategies*. Hodder Arnold, London, UK,
- Quiñones-Mateu ME, Avila S, Reyes-Teran G & Martinez MA (2014) Deep sequencing: Becoming a critical tool in clinical virology. *Journal of Clinical Virology* 61(1):9-19.
- Rao ST & Rossmann MG (1973) Comparison of super-secondary structures in proteins. *Journal of Molecular Biology* 76(2):241-256.
- Ren X, Hoiczyk E & Rasgon JL (2008) Viral paratransgenesis in the malaria vector *Anopheles gambiae*. *PLoS Pathogens* 4(8):e1000135.
- Reyes GR & Kim JP (1991) Sequence-independent, single-primer amplification (SISPA) of complex DNA populations. *Molecular and Cellular Probes* 5(6):473-481.
- Rhode S & Richard SM (1987) Characterization of the trans-activation-responsive element of the parvovirus H-1 P38 promoter. *Journal of Virology* 61(9):2807-2815.
- Ritchie BW, Niagro FD, Lukert PD, Steffens III WL & Latimer KS (1989) Characterization of a new virus from cockatoos with psittacine beak and feather disease. *Virology* 171(1):83-88.
- Roberts MM, White JL, Grutter MG & Burnett RM (1986) Three-dimensional structure of the adenovirus major coat protein hexon. *Science* 232(4754):1148-1151.
- Ros C, Gerber M & Kempf C (2006) Conformational changes in the VP1-unique region of native human parvovirus B19 lead to exposure of internal sequences that play a role in virus neutralization and infectivity. *Journal of Virology* 80(24):12017-12024.
- Rosario K, Duffy S & Breitbart M (2012) A field guide to eukaryotic circular single-stranded DNA viruses: insights gained from metagenomics. *Archives of Virology* 157(10):1851-1871.
- Rossmann MG (1990) The molecular replacement method. *Acta Crystallographica Section A: Foundations of Crystallography* 46(2):73-82.
- Rossmann MG, Arnold E, Erickson JW, Frankenberger EA, Griffith JP, Hecht H-J, Johnson JE, Kamer G, Luo M & Mosser AG (1985) Structure of a human common cold virus and functional relationship to other picornaviruses. *Nature* 317(6033):145-153.
- Rossmann MG, Bernal R & Pletnev SV (2001) Combining electron microscopic with X-ray crystallographic structures. *Journal of Structural Biology* 136(3):190-200.
- Rossmann MG & Johnson JE (1989) Icosahedral RNA virus structure. *ARB* 58(1):533-569.

- Santoro C, Mermod N, Andrews PC & Tjian R (1988) A family of human CCAAT-box-binding proteins active in transcription and DNA replication: cloning and expression of multiple cDNAs.
- Scheres SH & Chen S (2012) Prevention of overfitting in cryo-EM structure determination. *Nature Methods* 9(9):853-854.
- Schoborg RV & Pintel DJ (1991) Accumulation of MVM gene products is differentially regulated by transcription initiation, RNA processing and protein stability. *Virology* 181(1):22-34.
- Shimizu T (1975) Pathogenicity of an infectious flacherie virus of the silkworm, *Bombyx mori*, obtained from sericultural farms in the suburbs of Ina city [Japan]. *Journal of Sericultural Science of Japan*.
- Simpson AA, Chipman PR, Baker TS, Tijssen P & Rossmann MG (1998) The structure of an insect parvovirus (*Galleria mellonella* densovirus) at 3.7 Å resolution. *Structure* 6(11):1355-1367.
- Simpson AA, Hébert Bt, Sullivan GM, Parrish CR, Zádori Z, Tijssen P & Rossmann MG (2002) The structure of porcine parvovirus: comparison with related viruses. *Journal of Molecular Biology* 315(5):1189-1198.
- Styer E & Hamm J (1991) Report of a densovirus in a commercial cricket operation in the southeastern United States. *Journal of Invertebrate Pathology* 58(2):283-285.
- Sukhumsirichart W, Attasart P, Boonsaeng V & Panyim S (2006) Complete nucleotide sequence and genomic organization of hepatopancreatic parvovirus (HPV) of *Penaeus monodon*. *Virology* 346(2):266-277.
- Szelei J, Woodring J, Goettel M, Duke G, Jousset F-X, Liu K, Zadóri Z, Li Y, Styer E & Boucias D (2011) Susceptibility of North-American and European crickets to *Acheta domesticus* densovirus (AdDNV) and associated epizootics. *Journal of Invertebrate Pathology* 106(3):394-399.
- Tal J & Attathom T (1993) Insecticidal potential of the insect parvovirus GmDNV. *Archives of Insect Biochemistry and Physiology* 22(3 - 4):345-356.
- Tamura K, Dudley J, Nei M & Kumar S (2007) MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. *Molecular Biology and Evolution* 24(8):1596-1599.
- Tattersall P, Bergoin M, Bloom M, Brown K, Linden R, Muzychka N, Parrish C & Tijssen P (2005) Family *Parvoviridae*. *Virus taxonomy: VIIIth report of the International Committee on Taxonomy of Viruses*, Academic Press, San Diego, p 353-369.

- Thao ML, Wineriter S, Buckingham G & Baumann P (2001) Genetic characterization of a putative densovirus from the mealybug *Planococcus citri*. *Current Microbiology* 43(6):457-458.
- Tijssen P (1995) Densonucleosis viruses constitute an increasingly diversified subfamily among the parvoviruses. *Seminars in Virology*. Elsevier, p 347-355.
- Tijssen P (1999) Molecular and structural basis of the evolution of parvovirus tropism. *Acta Veterinaria Hungarica* 47(3):379-394.
- Tijssen P, Agbandje-McKenna M, Almendral J, Bergoin M, Flegel T, Hedman K, Kleinschmidt J, Li Y, Pintel D & Tattersall P (2011) *Virus taxonomy: classification and nomenclature of viruses: ninth report of the International Committee on Taxonomy of Viruses*. Elsevier, San Diego, CA.
- Tijssen P, Bando H, Li Y, Jousset F, Zadori Z, Fediere G, El-Far M, Szelei J & Bergoin M (2006a) Evolution of densoviruses. *Parvoviruses*, Jonathan K, Susan C & E BM (Édit.) CRC Press. p 55-68.
- Tijssen P & Kurstak E (1981) Biochemical, biophysical, and biological properties of densonucleosis virus (parvovirus). III. common sequences of structural proteins. *Journal of Virology* 37(1):17-23.
- Tijssen P, Li Y, El-Far M, Szelei J, Letarte M & Zádori Z (2003) Organization and expression strategy of the ambisense genome of densonucleosis virus of *Galleria mellonella*. *Journal of Virology* 77(19):10357-10365.
- Tijssen P, Szelei J & Zádori Z (2006b) Phospholipase A2 domains in structural proteins of parvoviruses. *Parvoviruses*, Jonathan K, Susan C & E BM (Édit.) CRC Press. p 95-105.
- Tijssen P, Tijssen-Van der Slikke T & Kurstak E (1977) Biochemical, biophysical, and biological properties of densonucleosis virus (paravovirus). II. Two types of infectious virions. *Journal of Virology* 21(1):225-231.
- Tijssen P, Van den Hurk J & Kurstak E (1976) Biochemical, biophysical, and biological properties of densonucleosis virus. I. Structural proteins. *Journal of Virology* 17(3):686-691.
- Tijssen P & Zádori Z (2002) Viral phospholipase A2 enzymes, anti-viral agents and methods of use. *WO patent 2,002,000,924*.
- Tischer I, Mields W, Wolff D, Vagt M & Griem W (1986) Studies on epidemiology and pathogenicity of porcine circovirus. *Archives of Virology* 91(3-4):271-276.
- Todd D, Niagro F, Ritchie B, Curran W, Allan G, Lukert P, Latimer K, Steffens III W & McNulty M (1991) Comparison of three animal viruses with circular single-stranded DNA genomes. *Archives of Virology* 117(1-2):129-135.

Tong L & Rossmann MG (1997) Rotation function calculations with GLRF program. *Methods in Enzymology* 276:594.

Towbin H, Staehelin T & Gordon J (1979) Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: procedure and some applications. *Proceedings of the National Academy of Sciences* 76(9):4350-4354.

Tsao J, Chapman MS, Agbandje M, Keller W, Smith K, Wu H, Luo M, Smith TJ, Rossmann MG & Compans RW (1991) The three-dimensional structure of canine parvovirus and its functional implications. *Science* 251(5000):1456-1464.

Tullis GE, Burger LR & Pintel DJ (1993) The minor capsid protein VP1 of the autonomous parvovirus minute virus of mice is dispensable for encapsidation of progeny single-stranded DNA but is required for infectivity. *Journal of Virology* 67(1):131-141.

Vago C, Duthoit J & Delahaye F (1966) Les lésions nucléaires de la virose à noyaux denses du lépidoptère *Galleria mellonella*. *Arch Gesamte Virusforsch* 18:344-349.

van Munster M, Dullemans A, Verbeek M, van den Heuvel J, Reinbold C, Brault V, Clerivet A & van der Wilk F (2003a) Characterization of a new densovirus infecting the green peach aphid *Myzus persicae*. *Journal of Invertebrate Pathology* 84(1):6-14.

van Munster M, Dullemans A, Verbeek M, van den Heuvel J, Reinbold C, Brault V, Clerivet A & Van der Wilk F (2003b) A new virus infecting *Myzus persicae* has a genome organization similar to the species of the genus *Densovirus*. *Journal of General Virology* 84(1):165-172.

Vaughn J, Goodwin R, Tompkins G & McCawley P (1977) The establishment of two cell lines from the insect *Spodoptera frugiperda* (Lepidoptera; Noctuidae). *In Vitro* 13(4):213-217.

Venkatakrishnan B, Yarbrough J, Domsic J, Bennett A, Bothner B, Kozyreva OG, Samulski RJ, Muzyczka N, McKenna R & Agbandje-McKenna M (2013) Structure and dynamics of adeno-associated virus serotype 1 VP1-unique N-terminal domain and its role in capsid trafficking. *Journal of Virology* 87(9):4974-4984.

Walters RW, Agbandje-McKenna M, Bowman VD, Moninger TO, Olson NH, Seiler M, Chiorini JA, Baker TS & Zabner J (2004) Structure of adeno-associated virus serotype 5. *Journal of Virology* 78(7):3361-3371.

Wang J, Zhang J, Jiang H, Liu C, Yi F & Hu Y (2005) Nucleotide sequence and genomic organization of a newly isolated densovirus infecting *Dendrolimus punctatus*. *Journal of General Virology* 86(8):2169-2173.

Wang Y, Abd-Alla A, Bossin H, Li Y & Bergoin M (2013) Analysis of the transcription strategy of the *Junonia coenia* densovirus (JcDNV) genome. *Virus Research* 174(1-2):101-107.

- Weissman DB, Gray DA, Pham HT & Tijssen P (2012) Billions and billions sold: pet-feeder crickets (Orthoptera: *Gryllidae*), commercial cricket farms, an epizootic densovirus, and government regulations make a potential disaster. *Zootaxa* 3504:67-88.
- Wen Z, Rupasinghe S, Niu G, Berenbaum MR & Schuler MA (2006) CYP6B1 and CYP6B3 of the black swallowtail (*Papilio polyxenes*): adaptive evolution through subfunctionalization. *Molecular Biology and Evolution* 23(12):2434-2443.
- Willwand K, Moroianu A, Hörlein R, Stremmel W & Rommelaere J (2002) Specific interaction of the nonstructural protein NS1 of minute virus of mice (MVM) with [ACCA] 2 motifs in the centre of the right-end MVM DNA palindrome induces hairpin-primed viral DNA replication. *Journal of General Virology* 83(7):1659-1664.
- Xiao C, Kuznetsov YG, Sun S, Hafenstein SL, Kostyuchenko VA, Chipman PR, Suzan-Monti M, Raoult D, McPherson A & Rossmann MG (2009) Structural studies of the giant mimivirus. *PLoS Biology* 7(4):e1000092.
- Xie Q, Bu W, Bhatia S, Hare J, Somasundaram T, Azzi A & Chapman MS (2002) The atomic structure of adeno-associated virus (AAV-2), a vector for human gene therapy. *Proceedings of the National Academy of Sciences* 99(16):10405-10410.
- Xie Q & Chapman MS (1996) Canine parvovirus capsid structure, analyzed at 2.9 Å resolution. *Journal of Molecular Biology* 264(3):497-520.
- Xu P, Cheng P, Liu Z, Li Y, Murphy RW & Wu K (2012) Complete Genome Sequence of a Monosense Densovirus Infecting the Cotton Bollworm, *Helicoverpa armigera*. *Journal of Virology* 86(19):10909-10909.
- Yamagishi J, Hu Y, Zheng J & Bando H (1999) Genome organization and mRNA structure of *Periplaneta fuliginosa* densovirus imply alternative splicing involvement in viral gene expression. *Archives of Virology* 144(11):2111-2124.
- Yu Q, Fadière G, Abd-Alla A, Bergoin M & Tijssen P (2012a) Iteravirus-like genome organization of a densovirus from *Sibine fusca* Stoll. *Journal of Virology* 86(16):8897-8898.
- Yu Q, Hajek AE, Bergoin M & Tijssen P (2012b) *Papilio polyxenes* densovirus has an iteravirus-like genome organization. *Journal of Virology* 86(17):9534-9535.
- Yu Q & Tijssen P (2014a) Iteradensovirus from the Monarch Butterfly, *Danaus plexippus plexippus*. *Genome Announcements* 2(2):e00321-00314.
- Yu Q & Tijssen P (2014b) Gene expression of five different iteradensoviruses: BmDV, CeDV, PpDV, SfDV and DpIDV. *Journal of Virology* 88:12152–12157
- Zádori Z, Szelei J, Lacoste M-C, Li Y, Gariépy S, Raymond P, Allaire M, Nabi IR & Tijssen P (2001) A Viral Phospholipase A₂ Is Required for Parvovirus Infectivity. *Developmental Cell* 1(2):291-302.

Zhai Y-g, Lv X-j, Sun X-h, Fu S-h, Fen Y, Tong S-x, Wang Z, Tang Q, Attoui H & Liang G-d (2008) Isolation and characterization of the full coding sequence of a novel densovirus from the mosquito *Culex pipiens pallens*. *Journal of General Virology* 89(1):195-199.

Zhi N, Zádori Z, Brown KE & Tijssen P (2004) Construction and sequencing of an infectious clone of the human parvovirus B19. *Virology* 318(1):142-152.

Appendix I: The Structure of Densovirus and Parvovirus

Article 7: The Structure and Host Entry of an Invertebrate Parvovirus

Geng Meng ^a, Xinzheng Zhang ^a, Pavel Plevka ^{a*}, Qian Yu ^b, Peter Tijssen ^b, Michael G. Rossmann ^a

Department of Biological Sciences, Purdue University, West Lafayette, Indiana, USA
^a; INRS-Institut Armand-Frappier, Université du Québec, Laval, Québec, Canada ^b

Contribution of authors

G. Meng, X. Zhang, P. Plevka, Q. Yu, P. Tijssen and M. G. Rossmann designed research; Q. Yu collected crickets, purified virus and prepared virus for analysis. G. Meng, X. Zhang and P. Plevka analyzed virus structure data; G. Meng, P. Tijssen, and M.G.Rossmann prepared the manuscript.

N.B. References are listed in the Reference section of the main body of the thesis (P. 154)

Abstract

The 3.5-Å resolution X-ray crystal structure of mature cricket parvovirus (*Acheta domesticus* densovirus [AdDNV]) has been determined. Structural comparisons show that vertebrate and invertebrate parvoviruses have evolved independently, although there are common structural features among all parvovirus capsid proteins. It was shown that raising the temperature of the AdDNV particles caused a loss of their genomes. The structure of these emptied particles was determined by cryo-electron microscopy to 5.5-Å resolution, and the capsid structure was found to be the same as that for the full, mature virus except for the absence of the three ordered nucleotides observed in the crystal structure. The viral protein 1 (VP1) amino termini could be externalized without significant damage to the capsid. *In vitro*, this externalization of the VP1 amino termini is accompanied by the release of the viral genome.

Résumé

La structure du parvovirus de criquet (*Acheta domesticus* densovirus [AdDNV]) a été déterminée par cristallographie aux rayons X avec une résolution de 3.5-Å. La comparaison des structures de parvovirus de vertébrés et d'invertébrés a montré que l'évolution de ces parvovirus a été effectuée indépendamment malgré certaines caractéristiques structurales communes à tous les parvovirus. Il a été montré que l'augmentation de la température de la capside du virus AdDNV a causé l'éjection du génome. La structure des capsides vides a été déterminée par cryo-microscopie électronique avec une résolution de 5.5-Å et la structure était la même que pour les particules virales matures complètes à l'exception de l'absence des trois nucléotides ordonnés observés dans la structure cristallisée. La région N-terminale de la protéine VP1 a pu être externalisée sans dommage significatif à la capside. *In vitro*, cette externalisation est accompagnée par la relâche du génome viral.

Introduction

Parvoviruses are small (250- to 300-Å-diameter), single-stranded DNA (ssDNA), icosahedral ($T = 1$), nonenveloped viruses whose genomes are approximately 5 kb long (Berns *et al.*, 2006). The *Parvoviridae* family has been subdivided into viruses that infect vertebrates (*Parvovirinae*) and those that infect invertebrates (*Densovirinae*). Parvoviruses replicate in dividing cells such as in tissues from insect larvae and fetuses. Densovirus are highly pathogenic, and those that use insect hosts usually kill 90% of the larvae within a few days (Bergoin *et al.*, 2000). Densovirus pose a threat to commercial invertebrates such as shrimp (Harvell *et al.*, 1999), silkworms (Li *et al.*, 2001), and crickets (Liu *et al.*, 2011, Szelei *et al.*, 2011). Some highly pathogenic densovirus are potential selective pesticides for vectors that transmit mosquito-borne diseases (Carlson *et al.*, 2006). *Parvovirinae* generally have three types of proteins (VP1, VP2, and VP3) in their capsids (Tijssen, 1999), whereas *Densovirinae* generally have four types of proteins (VP1 to VP4) in their capsids (Bergoin *et al.*, 2000). In densovirus there are 200 additional amino acids in VP1 at the N terminus. These different proteins result from different initiation sites for translation of the capsid gene and from post translational modification of their N termini (Tijssen, 1999). Generally, each of the 60 subunits within a capsid has the same amino acid sequence and is structurally the same, except that the different proteins start at different amino acids. The VP2s of some densovirus are unique among VP2s of parvoviruses since they are not completely contained within corresponding VP1s (Figure a1.1A).

Parvoviruses enter cells by dynamin-dependent receptor-mediated endocytosis and escape the endosome by the phospholipase (PLA2) activity within the amino-terminal domain of VP1 (Canaan *et al.*, 2004, Farr *et al.*, 2005, Girod *et al.*, 2002, Tijssen *et al.*, 2002, Zádori *et al.*, 2001). Although there is often less than 5% amino acid identity among the structural proteins of parvoviruses, the sequence of the PLA2 N-terminal domain of VP1 has more than 30% amino acid identity (Figure a1.1A and B). The PLA2 domain is not exposed in assembled, full parvoviruses such as minute

virus of mice (MVM) (Farr *et al.*, 2005) and human parvovirus B19 (Ros *et al.*, 2006), and it therefore has to be exposed during endocytosis (Canaan *et al.*, 2004, Farr *et al.*, 2005, Mani *et al.*, 2006, Ros *et al.*, 2006, Zádori *et al.*, 2001). However, the mechanism by which the VP1 amino-terminal PLA2 domain is exposed has not been elucidated in detail (Venkatakrishnan *et al.*, 2013).

The structures of six autonomous vertebrate parvoviruses (canine parvovirus [CPV] (Xie *et al.*, 1996), feline parvovirus [FPV]) (Agbandje *et al.*, 1993), porcine parvovirus [PPV] (Simpson *et al.*, 2002), MVM (Llamas-Saiz *et al.*, 1997), H-1 parvovirus [H-1PV] (Halder *et al.*, 2013), and human parvovirus B19 (Kaufmann *et al.*, 2004)) and three invertebrate parvoviruses (Galleria mellonella densovirus [GmDNV] (Simpson *et al.*, 1998), Bombyx mori densovirus [BmDNV] (Kaufmann *et al.*, 2011), and Penaeus stylirostris densovirus [PstDNV] (Kaufmann *et al.*, 2010)) have been determined (Table a1.1). Furthermore, extensive studies have been made of the human adeno-associated dependoviruses (Chapman *et al.*, 2006, Xie *et al.*, 2002). The structures of these parvoviruses consist of 60 structurally equivalent capsid proteins assembled with icosahedral symmetry. Each capsid protein has a “jelly roll” fold, a motif that is common to many viruses, including the nonenveloped RNA picornaviruses (Rossmann *et al.*, 1985) and small RNA plant viruses (Harrison *et al.*, 1978), as well as larger double-stranded DNA (dsDNA) adenoviruses (Roberts *et al.*, 1986), the enveloped bacteriophage PRD1 (Benson *et al.*, 1999), the fungal virus Paramecium bursaria chlorella virus 1 (PBCV-1) (Nandhagopal *et al.*, 2002), vaccinia virus (Bahar *et al.*, 2011), and probably also mimivirus (Xiao *et al.*, 2009). The jelly roll fold is a β -barrel consisting of two opposed antiparallel-sheets with adjacent-strand BIDG and CHEF, where the strands along the polypeptide chain are named A and B to H. The interior of the barrel is exceedingly hydrophobic.

Table a1.1 Structural studies of autonomous parvoviruses.

Virus	Description of particle	Structural protein(s) in particles	Resolution (Å)	Icosahedral ordered genome structure (bp)	PDB code (reference)
Vertebrate parvoviruses					
Canine parvovirus	Full virus Empty particle	VP1, VP2, VP3 VP1, VP2	2.9 3.0	11 None	4DPV (17) 2CAS (56)
Feline parvovirus	Empty particle	VP3	3.3	None	1FPV (18)
Porcine parvovirus	Virus-like particle	VP2	3.5	None	1K3V (19)
Human parvovirus B19	Virus-like particle	VP2	3.5	None	1S58 (22)
Minute virus of mice	Full virus	VP1, VP2, VP3	3.5	11	1MVM (20)
Rat H-1 parvovirus	Full virus Empty particle	VP1, VP2, VP3 Unknown	2.7 3.2	10 None	4G0R (21) 4GBT
Invertebrate parvoviruses					
<i>Gm</i> DNV	Full virus	VP1, VP2, VP3, VP4	3.7	None	1DNV (23)
<i>Bm</i> DNV	Virus-like particle	VP3	3.1	None	3P0S (24)
<i>Pst</i> DNV	Virus-like particle	VP4	2.5	None	3N7X (25)
<i>Ad</i> DNV	Full virus Induced emptied particle	VP1, VP2, VP3, VP4 VP1, VP2, VP3, VP4	3.5 5.5	3 None	

Parvoviruses have a channel along the 5-fold axes formed by five symmetry-related DE loops (the “DE” loop is between the β-strands D and E). Residues lining the channel are mostly hydrophobic and guide the externalization of a conserved glycine-rich sequence near the amino ends of the VPs (Agbandje *et al.*, 1998, Chapman *et al.*, 1993, Tsao *et al.*, 1991). The loops connecting the β-strands of the jelly roll fold are usually exceptionally large in parvoviruses compared with the loops in picornaviruses (Hogle *et al.*, 1985, Rossmann *et al.*, 1985) and form the exterior of the virus and intersubunit contacts (Figure a1.1C). These loops are more variable in sequence than the core jelly roll structure.

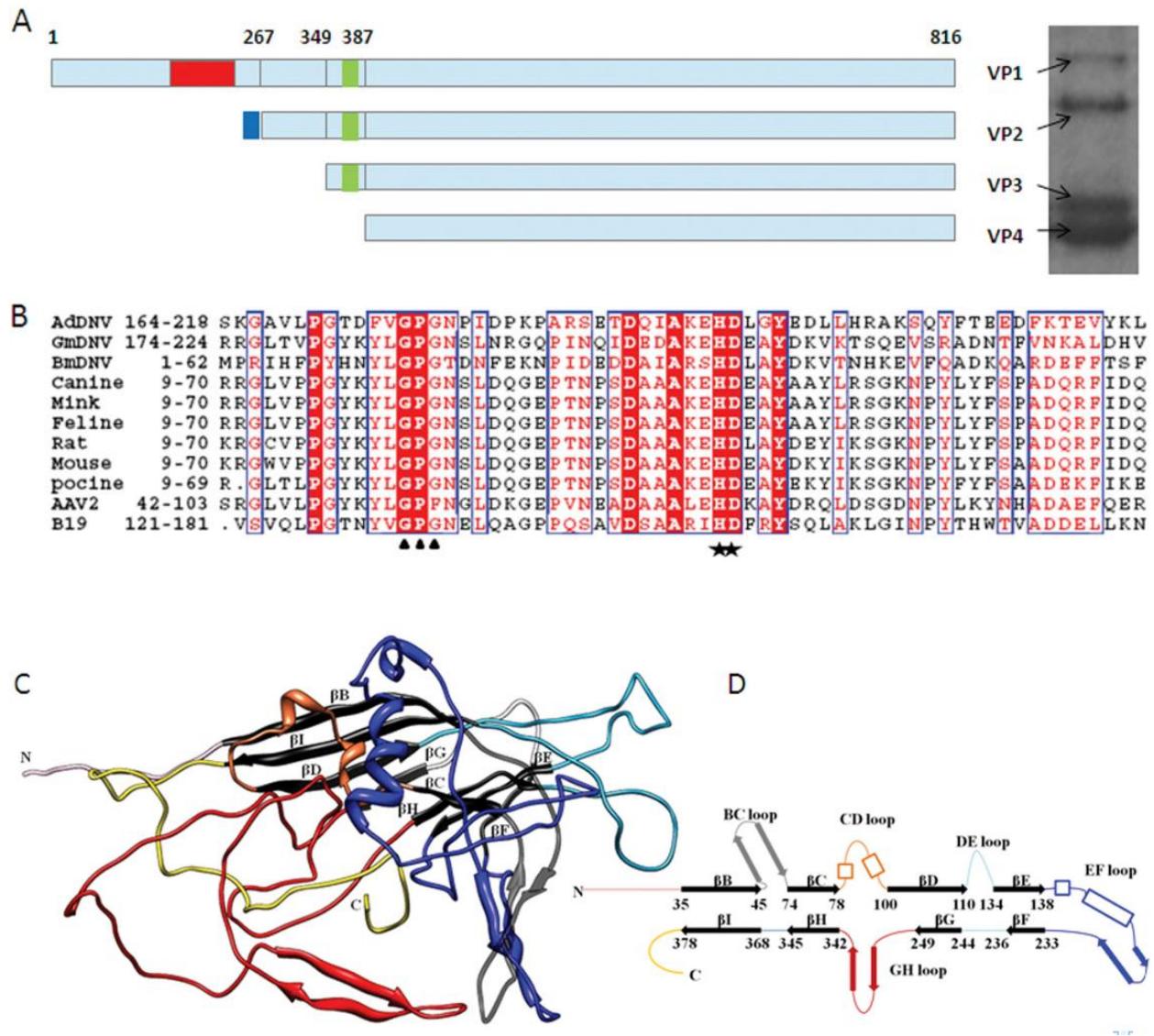


Figure a1.1 Structure of the AdDNV capsid protein.

(A) Left, capsid protein organization. Red area, PLA2 domain; green area, glycine-rich region; blue area, extra sequence at the N termini of VP2. Right, SDS-polyacrylamide gel of the AdDNV full particles. (B) Alignment of PLA2 sequences in VP1s of various parvoviruses, including AdDNV, GmDNV, BmDNV, CPV, mink parvovirus, FDV, rat parvovirus, mouse parvovirus, PPV, AAV2, and B19. The stars indicate the His-As catalytic site, and the triangles locate the Ca²⁺-binding site. (C) Three-dimensional structure of AdDNV capsid protein, showing the core jelly roll in black. The surface loops connecting the strands of the core jelly roll are colored as follows: BC loop, gray; CD loop, orange; DE loop, sky blue; EF loop, dark blue; and GH loop, red. (D) Diagrammatic representation of the capsid protein structure (heavy lines represent β-strands) (color coding is the same as in panel C).

Here, we describe the crystal structure of mature virions of cricket parvovirus (*Acheta domesticus* densovirus [AdDNV]) at 3.5-Å resolution and the cryo-electron microscopic (cryoEM) structure of the emptied virus at 5.5-Å resolution. We also report on the externalization of the VP1 N-terminal region and subsequent genome release by an increase in temperature.

MATERIALS AND METHODS

Virus purification and preparation of the emptied virus particles.

The original virus was isolated from infected crickets (Liu *et al.*, 2011). Further purification was achieved using CsCl equilibrium density gradient centrifugation. Of the two bands with different densities, the lower band contained the full particles and represented 99% of all the particles. The upper band contained empty particles assembled mainly from VP4. The two bands were separately transferred into Tris-buffered saline (TBS) (10 mM Tris-Cl, 100 mM NaCl, 1 mM CaCl, and 1 mM MgCl at pH 7.5) for further usage. Aliquots of the full virus particles were incubated at 26, 37, 45, 55, 65, 75, and 100°C for 1 h. The heat-treated emptied particles were frozen on holey carbon (Quantifoil) EM grids and checked by cryoEM. The numbers of full, empty, and broken particles were counted by eye (Figure a1.2) and averaged over three holes on two different EM grids. Each hole had roughly 100 particles.

Determination of the crystal structure of the full virus particle.

Crystals of the full particles were obtained by hanging-drop vapor diffusion in the presence of 20% polyethylene glycol (PEG) 400 and 100 mM MgCl₂ at 16°C. Further optimization of the crystallization conditions produced crystals of up to 0.5 mm in length. Crystals were soaked for at least 20 min in the presence of 20% glycerol cryoprotectant prior to freezing.

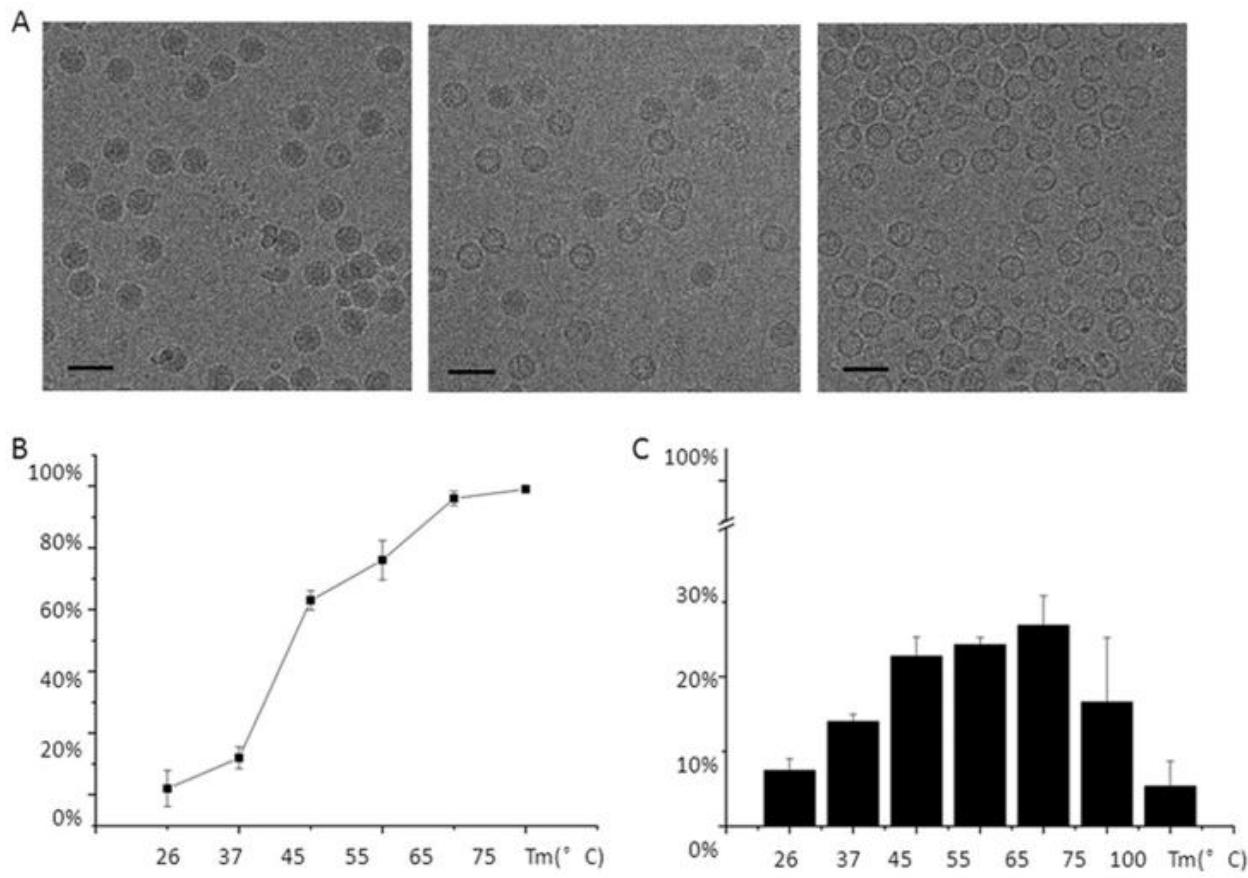


Figure a1.2 Heat treatment and PLA2 activity of the virus particles.

(A) CryoEM micrographs showing the virus after incubation for 1 h at 26°C, 45°C, and 65°C, resulting in only full particles (left), about an equal number of full and emptied particles (middle), and emptied particles (right), respectively. (B) Percentage of emptied particles after incubating full particles at different temperatures. (C) Phospholipase activity as a function of temperature with respect to the PLA2 activity of honey bee PLA2 at 26°C.

X-ray diffraction data were collected at 100 K at the Advanced Photon Source (APS) beamline 23ID (Table a1.2). Diffraction data from about 20 crystals were indexed and merged, using the HKL2000 computer program (Otwinowski *et al.*, 1997) to generate the final 3.5-Å resolution data set. The space group was P4 2 2 1 2 with a 412.67 Å and c 278.80 Å. The Matthews coefficient was 3.64 Å³/Da, assuming half a virus particle per crystallographic asymmetric unit. Thus, the virus was located on a crystallographic 2-fold axis. A self-rotation function, calculated with the GLRF program (Tong *et al.*, 1997) using 8- to 3.5-Å resolution data, gave the accurate orientation of the particle about the crystallographic 2-fold axis. This showed that one of the

icosahedral 2-fold axes of the virus was roughly parallel to the 4_2 crystallographic axis, with a 1.6° rotation away from being exactly parallel. As a consequence, the position of the particle along the crystallographic 2-fold axis could be determined from the big Patterson peak generated by the large number of parallel equal-length vectors.

Table a1.2 X-ray data collection and structure refinement.

Parameter	Value ^a
Data collection parameters	
Wavelength (Å)	1.0715
CCD detector	MAR CCD-325
Exposure time/frame (s)	3
Oscillation angle/frame ($^\circ$)	0.3
No. of frames collected	300
Data reduction and refinement statistics	
Resolution range (Å)	50–3.5 (3.6–3.5)
Space group	P4 ₂ 2 ₁ 2
No. of frames used	80
Cell parameters	
a, b, c (Å)	412.67, 412.67, 278.80
α, β, γ ($^\circ$)	90, 90, 90
Mosaicity ($^\circ$)	0.40
No. of observed reflections	1,518,629
No. of unique reflections	236,217 (9,755)
Redundancy	1.2 (1.1)
% Completeness	56 (46.5)
$\langle I \rangle / \langle \sigma(I) \rangle$	4.0 (1.2)
Rsym $\sum h \sum j I_{hj} - \langle I_h \rangle / \sum \sum I_{hj}$	0.161 (0.565)
Model building and refinement statistics	
Resolution range (Å)	30–3.5 (3.6–3.5)
No. of residues/atoms built	412/3,290
Final <i>R</i> factor ^b	0.289
Mean isotropic temp factor (Å ²)	18
RMSD bond length (Å)/bond angle ($^\circ$)	0.0047/1.258
% Residues in most favored/additionally allowed/generously allowed/disallowed regions of the Ramachandran plot ^c	82.5/16.9/0.3/0.3

The structure was determined using the molecular replacement method (Rossmann, 1990) with the structure of GmDNV (Protein Data Bank [PDB] code 1DNV) (Simpson *et al.*, 1998) as the initial phasing model to 15-Å resolution. The phases were then extended to 3.5-Å resolution in steps of one reciprocal lattice interval (1/c) at a time. Three cycles of 30-fold noncrystallographic symmetry (NCS) averaging and solvent flattening were performed for each extension step. The averaging and extension processes were performed using the program AVE in the Uppsala Software Factory (Kleywegt *et al.*, 1997) and FFT, SFALL in CCP4 programs (Collaborative, 1994). The final overall correlation coefficient between the observed structure amplitudes and the calculated structure factors corresponding to the final averaged and solvent-flattened map was 0.866. The atomic model was built into the 3.5-Å resolution map using COOT (Emsley *et al.*, 2004). The model coordinates were refined with the CNS program (Brunger *et al.*, 1998) while applying NCS constraints and reasonable model restraints, including group temperature factor refinement. No attempt was made to identify water molecules, as the data extended to only 3.5-Å resolution. The structures of the three ordered nucleotides bound to the inside surface of the capsid (see Results and Discussion) were included in the final stages of refinement. After more than 5 cycles of refinement and model rebuilding, the R factor had dropped from 34% to 28.9%. In the presence of the 30-fold NCS redundancy, there will be no significant difference between R free and R working.

Detecting the externalized VP1 N termini and their phospholipase A2 activity.

Full particles and the heat-treated emptied particles were digested by trypsin at room temperature. About 30 µl of the particle suspension at a concentration of 5 µg/µl was incubated with trypsin for 1h at 26°C. The trypsin had a final concentration of 1 µg/µl in the mixture. The samples were then checked for VP1 cleavage using a 15% SDS-polyacrylamide gel (Figure a1.3).

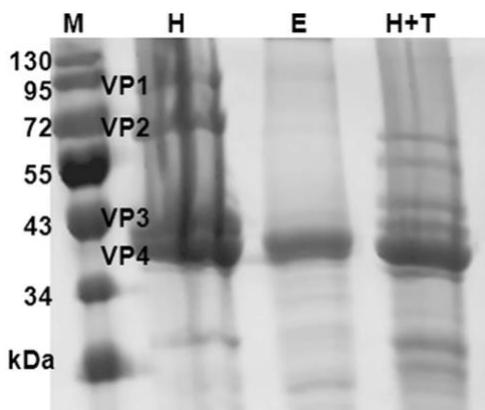


Figure a1.3 SDS-PAGE of heat-treated AdDNV particles.

Lanes: M, protein markers; H, heat-treated emptied AdDNV particles; E, non-heat-treated empty particles among the purified mature AdDNV particles; H+T, further trypsin digestion of the heat-treated emptied AdDNV particles.

The PLA2 activities of the full particles and the heat-treated emptied particles were measured at 26°C by a colorimetric assay (sPLA2 assay kit; Cayman Chemical, Ann Arbor, MI), using the 1,2-dithio analog of diheptanoyl phosphatidylcholine (diheptanoyl thio-PC) as the substrate for PLA2. The absorbance at 405 nm was determined every minute for 30 min. Measurements of the PLA2 activity were normalized relative to the activity of 1 ng bee venom PLA2. The activity of the PLA2 in 1.5 µg virus is equivalent to the PLA2 activity in 1 ng bee venom.

CryoEM and three-dimensional structural reconstruction of emptied particles.

The optimal condition for obtaining the largest percentage of emptied, unbroken particles was 55 °C for 1 h (Figure a1.2). Three microliters of the heat-treated emptied particles at a protein concentration of 5 µg/µl was applied to holey grids (Quantifoil) and blotted for 6 s in an FEI Mark 3 Vitrobot chamber at 90% humidity. The grids were then fast-frozen in liquid ethane. Cryo-electron microscopy (CryoEM) images were acquired on an FEI Titan Krios operated at 300 keV. Images were recorded with a 4k 4k charge-coupled device (CCD) detector. As a control, grids of untreated particles were prepared and viewed in the same way. The assumed magnification of 59,000 was calibrated with respect to a known specimen and was shown to correspond to a

pixel separation of 1.51 Å in the image. The electron dose was 20 e/Å², and the image was defocused by between 1.6 and 2.6 m. About 150 cryoEM micrographs, each showing roughly 100 particles, of the emptied particles were recorded. The defocus and the astigmatism of each micrograph were estimated with the EMAN1 fitctf program (Ludtke *et al.*, 1999) and further confirmed with the program ctfit. Image processing and three-dimensional reconstruction were performed using the EMAN suite of programs (Baker *et al.*, 2010). The final reconstruction was computed using 15,000 particles out of about 17,000 initial boxed images and was found to have 5.5-Å resolution based on the separate structure determinations of two randomly selected independent sets of images using the Fourier shell correlation threshold of 0.143 (Figure a1.4) (Scheres *et al.*, 2012).

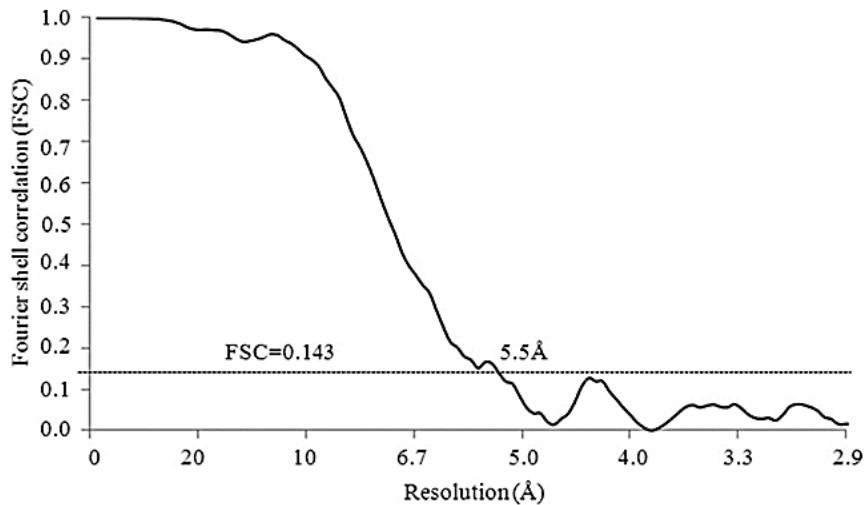


Figure a1.4 Fourier shell correlation (FSC).
FSC based on the independent structure determinations of two randomly selected equally sized sets of images, showing the resolution of the emptied AdDNV particle reconstruction to be ~5.5 Å when the FSC is 0.143.

Sequence alignment of the PLA2 domain and structural comparisons.

The sequence of the AdDNV VP1 N-terminal PLA2 domain (GI 326392953) was aligned with the corresponding sequences of adeno-associated virus 2 (AAV2) (GI 110645923), human parvovirus B19 (GI 169212578), CPV (GI 116646110), MVM (GI 332290), rat parvovirus (GI 410443463), mink parvovirus (GI 425696394), PPV (GI

46404508), GmDNV (GI 23334609), and BmDNV (GI 18025360) using Clustal X (Higgins *et al.*, 1988).

The crystal structure of AdDNV was compared with those of other invertebrate densoviruses, i.e., GmDNV (Simpson *et al.*, 1998) and BmDNV (Kaufmann *et al.*, 2011), as well as mammalian autonomous parvoviruses CPV (Xie *et al.*, 1996), PPV (Simpson *et al.*, 2002), FPV (Agbandje *et al.*, 1993), MVM (Llamas-Saiz *et al.*, 1997), and B19 (Kaufmann *et al.*, 2004) using the HOMOlogy program (Rao *et al.*, 1973). These structural comparisons do not include the disordered PLA2 domain, whose positions in the virus are random and therefore cannot be observed in the crystal structure.

Accession numbers. The atomic coordinates of the AdDNV crystal structures have been deposited with the Protein Data Bank (www.pdb.org) (PDB code 4MGU); the cryo-EM maps of the emptied AdDNV particle have been deposited with the Electron Microscopy Data Bank (www.emdatabank.org) (EMDB code EMD-2401).

RESULTS AND DISCUSSION

Crystal structure of the full AdDNV particles.

The structure of AdDNV was determined to 3.5-Å resolution. The position of the core jelly roll relative to the icosahedral symmetry axes was essentially the same in AdDNV as in other known parvovirus structures (Table a1.3 and a1.4).

Table a1.3 Sequence and structural comparisons of AdDNV capsid protein with other autonomous parvovirus capsid proteins.

Virus	Sequence identity (%)	RMSD (Å)		
		between C α atoms	No. of aligned C α atoms	Total no. of C α atoms
Canine parvovirus	<5	4.8	261	548
Feline parvovirus	<5	4.9	263	534
Porcine parvovirus	<5	4.9	258	542
Human B19	<5	5.1	260	523
Minute virus of mice	<5	5.0	264	549
<i>Gm</i> DNV	30.50	2.1	295	415
<i>Bm</i> DNV	<5	3.7	331	412
<i>Pst</i> DNV	8.30	4.1	224	299

Table a1.4 Superposition of the AdDNV jelly roll core (70 residues) on other invertebrate parvovirus capsid proteins.

Virus	RMSD (Å) between C α atoms
Canine parvovirus	2.0
<i>Gm</i> DNV	0.8
<i>Bm</i> DNV	1.4
<i>Pst</i> DNV	1.6

The four structural proteins VP1 (88.1 kDa), VP2 (65.3 kDa), VP3 (50.8 kDa), and VP4 (46.9 kDa) are in an approximate 1:11:18:30 proportion in AdDNV full particles based on scanning the gel with Kodak Image Station 2000R and analyzing with software Kodak MI (Figure a1.1A). The glycine-rich sequence is present in VP1, VP2, and VP3, but is missing in VP4 (Figure a1.1A). It may be significant that, compared with vertebrate parvoviruses, there is therefore only one copy of the PLA2 structure per virion. The polypeptide chain of the capsid protein could be traced from

residue 23 of VP4 situated at the base of the 5-fold axis channel to residue 418 at the carboxy terminus (Figure a1.1C).

The electron density in this channel (1.5) of AdDNV is weak and discontinuous (Figure a1.5C), which is similar to the density in the GmDNV 5-fold channel. The glycine-rich motif in AdDNV consists of about 17 residues, 8 of which are glycines, whereas in GmDNV the same motif has 7 glycines and about 16 residues (Figure a1.6). The difference of the sequence length may be partly related to the structure of the channel in the different parvoviruses.

The low density in the 5-fold channel suggests that only several of the 12 5-fold channels are occupied, resulting in externalization of the VP amino termini. A similar lack of amino-terminal externalization was observed in GmDNV, the only other known structure of a mature DNV. The structures of silkworm and shrimp densoviruses (Table a1.1) were self-assembled from recombinantly expressed VP3 and VP4 capsid proteins, respectively. Hence, these structures are missing the glycine-rich sequence. As there is only one VP1 per virion, some of the 5-fold channels must be occupied by VP2 or VP3. However, in the vertebrate parvoviruses CPV (Xie *et al.*, 1996) and MVM (Llamas-Saiz *et al.*, 1997), all the 5-fold channels are always fully occupied by the VP amino termini. Furthermore, there is a larger proportion of VP1 subunits per virion in the vertebrate parvoviruses. However, the externalized amino termini are disordered in both the vertebrate and invertebrate parvovirus crystal structures.

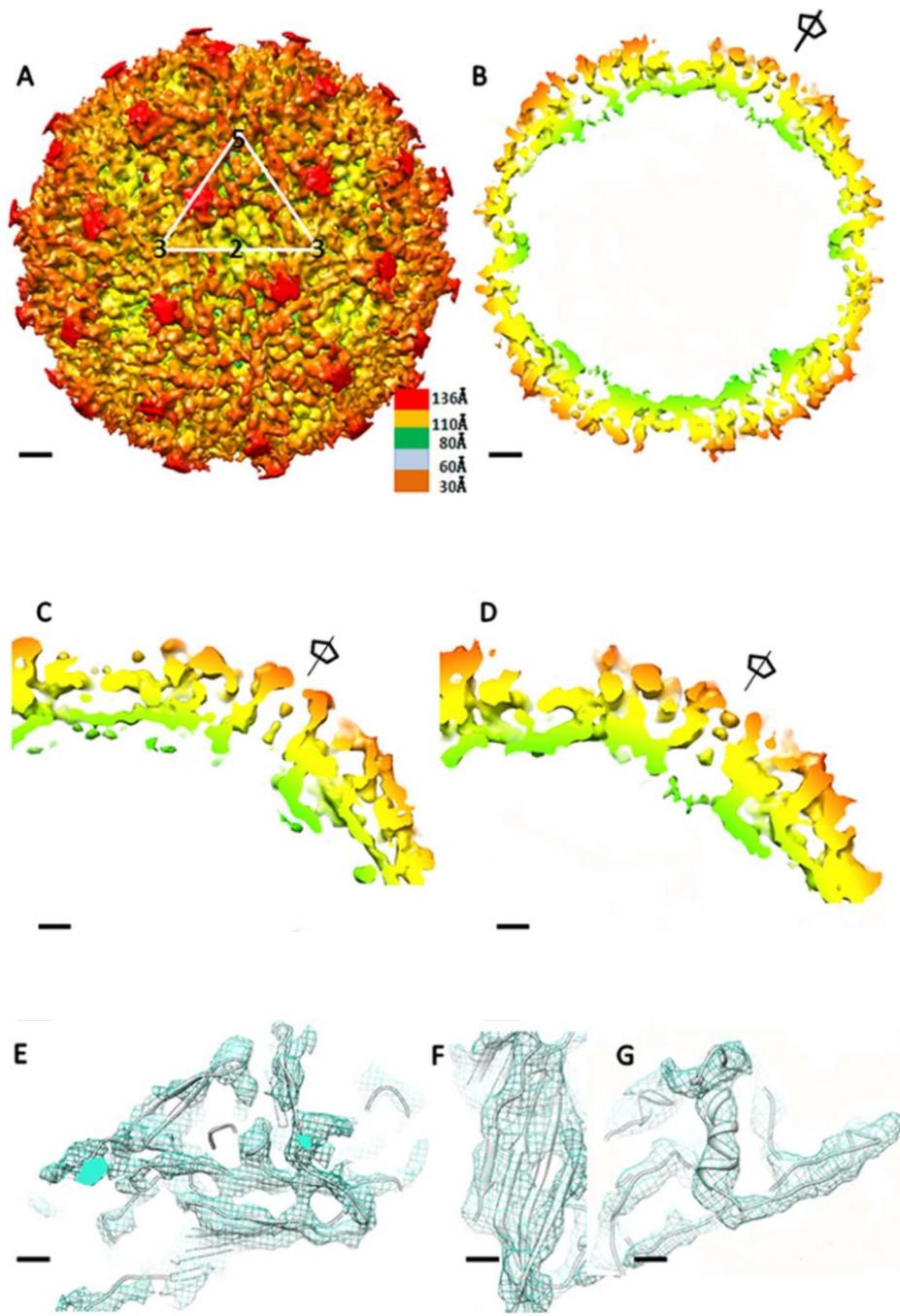


Figure a1.5 Structure of AdDNV emptied particles.

(A) CryoEM reconstruction of emptied particles. Surface features with a triangle showing the limits of one icosahedral asymmetric unit are shown. The scale bars represent 2 nm. (B) Center section of the cryoEM reconstruction. The scale bars represent 2 nm. (C) Enlargement of the 5-fold channel density in the X-ray electron density map. The scale bars represent 1 nm. (D) Enlargement of the 5-fold channel density in the cryoEM density map. The scale bars represent 1 nm. (E, F, and G) Fit of the X-ray structure polypeptide backbone into the cryoEM density for the β -sheets of the jelly roll (scale bars represent 5 Å) (E), the BIDG β -sheet (the scale bars represent 3 Å) (F), and the β -helix located in the EF loop (the scale bars represent 3 Å) (G).

AdDNVGKGKR	G	GG	RPPKSS	S	GG
GmDNAGTGS	G	T	SS	GGGNT	Q
BmDNA	GGGAQVDPR..TG.	G	Q	A	GG	GG
canine	GGQPAVRNERATGSGN	G	GG	GGGGG	S	GGVG.....
mink	GGQPAVRNERATGSGN	G	GG	GGGGG	S	GGVG.....
feline	GGQPAVRNERATGSGN	G	GG	GGGGG	S	GGVG.....
porcineGNESGG	G	GG	GGGRG	A	GGVGVSTG
B19GAGG	G	GS	NPVKSMWSEGAT..		

Figure a1.6 Glycine-rich regions comparison.

Sequence comparisons of the glycine-rich regions of AdDNV and other autonomous parvovirus, including GmDNA (GI 23334609), BmDNA (GI 18025360), canine parvovirus (CPV) (GI 116646110), mink enteritis virus (MEV) (GI 425696394), feline parvovirus (FPV) (GI 333476), porcine parvovirus (PPV), and human parvovirus (B19) (GI 169212578).

Pairwise comparisons of the parvovirus structures were used to determine the number of and root mean square deviation (RMSD) between equivalent C atoms (Table a1.3). The number of inserted amino acids between β -strands was used to calculate a phylogenetic tree using the program MEGA (Tamura *et al.*, 2007) (Figure a1.7). This shows a closer relationship among the insect densoviruses (AdDNV, GmDNA, PstDNA, and BmDNA) than between these viruses and vertebrate viruses. Although the capsid proteins of all parvoviruses have common structural features, the capsid proteins of vertebrate and invertebrate parvoviruses must have evolved independently (Figure a1.7).

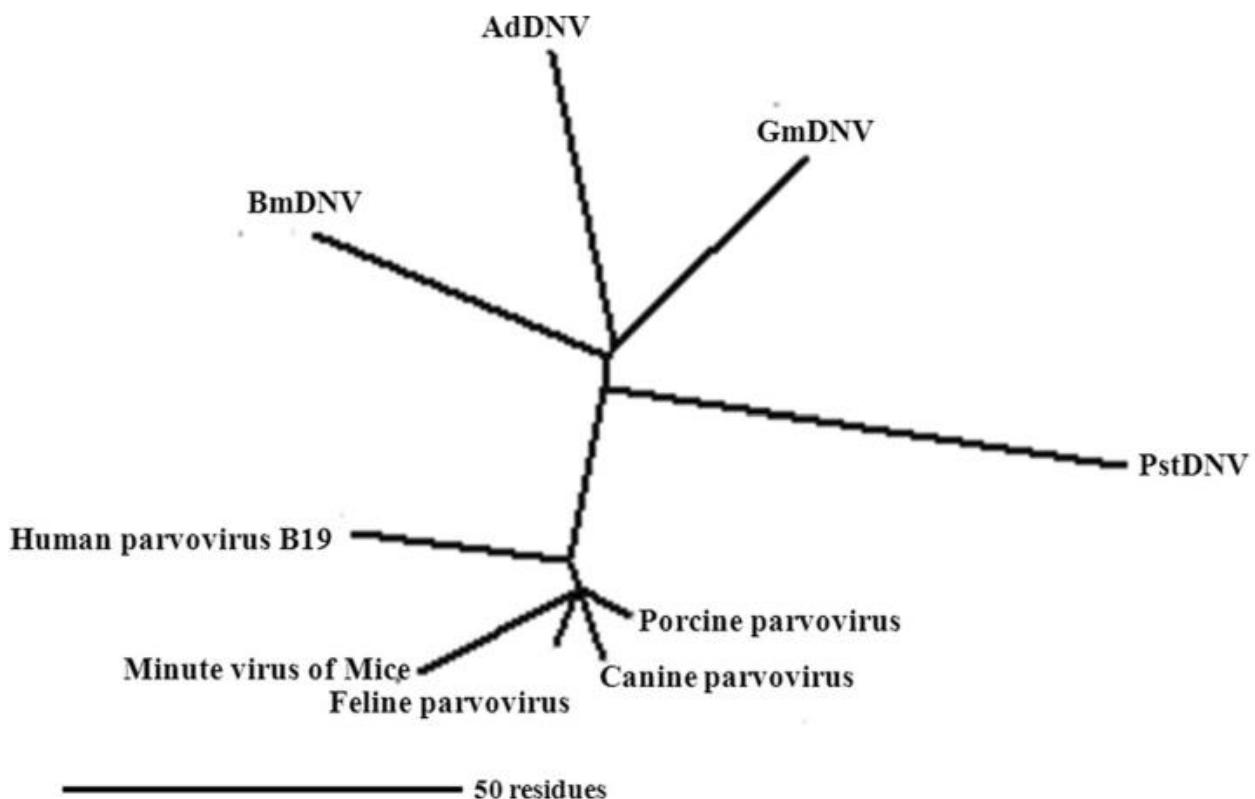


Figure a1.7 Phylogenetic analysis.

Unrooted phylogenetic tree of different parvoviruses based on the number of inserted residues between β -strands.

The A β -strand is folded back to run antiparallel to β -strand B in the BIDG sheet in all known vertebrate parvovirus structures (Figure a1.1D). In previously determined densovirus structures, including the AdDNV structure reported here, β -strand A is associated with β -strand B in the neighboring, 2-fold related BIDG sheet. Such an exchange of domains between 2-fold-related structures is an example of “domain swapping” (Figure a1.1C and D). However, in AdDNV the A β -strand contains three proline residues (Pro24, Pro26, and Pro28) and therefore diminishes the H bonding with the B β -strand of the neighboring subunit.

The channel along the 5-fold icosahedral axes of parvoviruses is formed by the DE loop and is between 16 Å and 18 Å in diameter, measured from atom center to atom center. The conformation of the DE loop is variable among both vertebrate and invertebrate parvoviruses (Tijssen, 1999). In AdDNV, as also in all other known parvoviruses, there are several hydrophobic amino acids in the DE loop (from 116 Ala to 133 Gln in AdDNV) that interact with the glycine-rich region.

Emptied densovirus particles and phospholipase activity.

Most parvoviruses, including densoviruses, assemble *in vivo* both as full infectious particles and as empty particles. However, for AdDNV and presumably also for other densoviruses, the small fraction of particles that are empty in a virus preparation consist of only VP4 (Figure a1.3) and are missing the glycine-rich sequence, whereas the dominant infectious virus particles contain all four types of subunits (VP1 to VP4) (Figure a1.1A). Therefore, after heat treatment, nearly all the emptied particles that have a full complement of all four VPs must have been full of genome, whereas empty particles containing only VP4 must have been assembled as empty particles. It had been shown that heating parvoviruses to 70°C generated PLA2 activity, suggesting exposure of the VP1 N termini (Farr *et al.*, 2005, Tijssen *et al.*, 2002). However, it was not clear whether only the VP1 N termini were exposed from intact particles or whether the particles had disassembled. The loss of the genome associated with a presumably transient change in the capsid has some resemblance to the infectious process in picornaviruses (Garriga *et al.*, 2012).

Here we used cryoEM to show that on heating of AdDNV for a defined length of time, the number of emptied particles increased with temperature (Figure a1.2A and B). When the temperature was increased beyond 65°C there was also an increase of broken particles. Concomitant with the increase of emptied particles, there was also an increase of PLA2 activity (Figure a1.2C). Above about 65°C, the virions disintegrated and had reduced PLA2 activity. Unlike the case for full, infectious AdDNV particles, the VP1 N termini of the heat-treated emptied particles were sensitive to

trypsin digestion, whereas the capsids remained intact as determined with cryoEM. This showed that the heat treatment causes externalization of the N termini while leaving the capsid intact, with the PLA2 domain remaining a part of the particle.

Externalization of the N-termini prior to endocytosis abolishes infectivity in parvovirus (Boisvert *et al.*, 2010), suggesting that the sequence of events during infection is critical. This could explain why all parvoviruses harbor the N-terminal part of VP1 within the virus particle until they are ready to breach the endosomal membrane. PLA2 requires a Ca^2 concentration of greater than 1 mM (Canaan *et al.*, 2004). Such Ca^2 concentrations are present in the endosome but not in the cytoplasm (Huotari *et al.*, 2011), further narrowing the viral PLA2 activity to the endosomal membrane.

The 5.5-Å pseudo-atomic-resolution cryoEM structure of emptied AdDNV particles.

An initial effort to crystallize heat-treated emptied particles failed to produce crystals, probably because of the externalized VP1 N termini. However, it was possible to obtain a 5.5-Å resolution cryoEM structure from approximately 15,000 images of heat-treated emptied particles (Figure a1.5A and B). The structure of the crystallographically determined virus could readily be superimposed onto the cryoEM map by aligning the icosahedral symmetry axes (Table a1.5). The cryoEM map had to be expanded by 5%, amounting to a radius increase of 6 Å, to obtain the best fit. This change in size of the EM map is well inside the error of determining the magnification of the electron microscope. The quality of the cryoEM map was excellent, as indicated by the resolution of the main chain in some places (Figure a1.5E to G).

Table a1.5 Results of a six-dimensional search on fitting of the AdDNV capsid protein structure into the 5.5-Å cryoEM density by using the EMfit program (55)^a.

Sumf ^b	Clash (%) ^c	-Den (%) ^d
40.0	1.7	4.8
24.0	14.6	15.6
21.1	19.4	20.9

^a Three possible fits were found, but the top fit is by far the best.

^b Sumf, mean density height averaged over all atoms, where the maximum density in the electron density map is set to 100.

^c Clash, percentage of atoms in the model that approach closer than 5 Å to icosahedral-related capsid protein molecules. ^d Den, percentage of atoms in density less than zero density.

A 5.5-Å resolution difference map between the crystal structure electron density and the cryoEM map was calculated using the EMfit program (Rossmann *et al.*, 2001). This showed two regions of density higher than three standard deviations of the background density, associated with the inside surface of the protein shell. The first region had an average height of about 6.5 Å and could be readily interpreted in terms of three nucleotides (Figure a1.8A). This structure was an extended trinucleotide with the bases facing the inside capsid surface (Figure a1.8B), close to the icosahedral 3-fold axes with the interaction with Tyr337 and Gln252. The second region had an average height of 3.5 but was not easily interpreted in terms of a standard nucleotide structure. Icosahedrally ordered genome structure has been previously observed in canine parvovirus (Tsao *et al.*, 1991) and minute virus of mice (Agbandje *et al.*, 1998) but not in invertebrate densoviruses (Table a1.1).

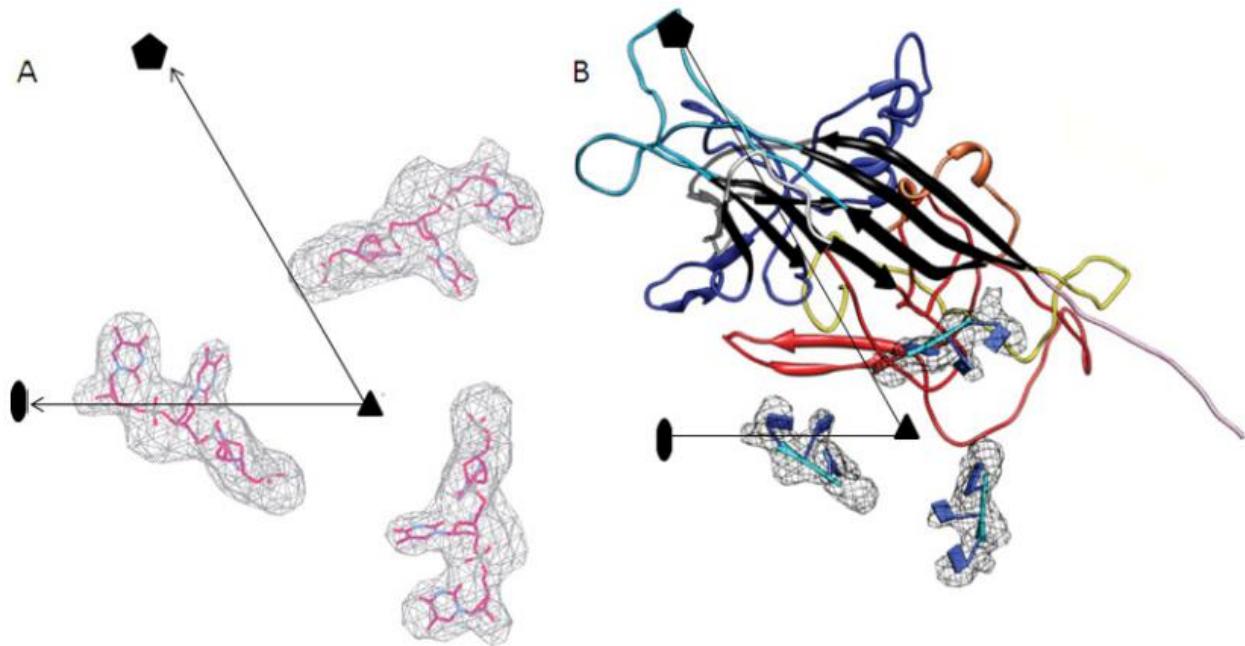


Figure a1.8 Structure of the three ordered ssDNA bases in AdDNV particles.
(A) Difference map between X-ray electron density and cryoEM density, with black lines showing the limits of one icosahedral asymmetric unit. Three bases of the ssDNA were built into the density. **(B)** Location of the three bases relative to the AdDNV capsid. The colors of the AdDNV capsid polypeptide are as defined for Fig. 1C.

The density in the channel along the 5-fold axes in the cryoEM map of the emptied particles was similar to that in the crystallographically determined map of the full infectious particle calculated to 5.5-Å resolution (Figure a1.5C and D). Thus, the emptied particles still have the glycine-rich region occupying the channel along the 5-fold axes in at least some of the 12 channels of each particle. The externalization of the VP1 termini does not seem to have caused much damage to the particles. In contrast to the case for AdDNV, there is little density along the 5-fold axes of the shrimp (PstDNV) and silkworm (BmDNV) densoviruses. However, these structures are of recombinant particles that contain only VP3 or VP4, respectively, and are therefore missing the glycine-rich region. These observations pose the intriguing question of whether the PLA2 domain has to be refolded to be threaded through the 5-fold pore or whether the pore opens with restoration of the initial capsid structure after extrusion.

ACKNOWLEDGMENTS

We thank Agustin Avila-Sakar and Valorie Bowman for help with the cryo-electron microscopy. We are grateful to Sheryl Kelly for help in preparing the manuscript. We also thank the staff of beamline 23ID, GM/CA-CAT, at Advanced Photon Source, Argonne National Laboratory, for help with the data collection.

Use of the Advanced Photon Source was supported by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences, under contract DE-AC02-06CH11357. The work was supported by an NIH grant award (AI11219) to M.G.R. The work was also supported by Purdue University funds for Structural Biology and the Electron Microscope Facility. P.T. was supported by a grant from the Natural Sciences and Engineering Research Council of Canada, and Q.Y. acknowledges tuition waivers at INRS-Institut Armand-Frappier and a scholarship from the People's Republic of China.

G.M., X.Z., P.P, Q.Y., P.T, and M.G.R. designed research; G.M., X.Z., P.P., and Q.Y. performed research; G.M. analyzed data; Q.Y. and P.T. contributed reagents/analytic tools; and G.M., P.T., and M.G.R. wrote the paper.

We declare no conflict of interest.

Article 8: The Structure of Serpine AAV Conserves the Capsid Topology of the Vertebrate AAVs

Paul Chipman¹, Kim Van Vliet¹, Robert McKenna¹, Qian Yu², Maude Boisvert², Peter Tijssen², Mavis Agbandje-McKenna¹

1 University of Florida, Department of Biochemistry & Molecular Biology, Gainesville, FL;

2 NRS-Institut ArmandFrappier, Université du Québec, Laval, QC, H7V 1B7, Canada.

Contribution of authors

P. Chipman, K.V. Vliet, R. Mckenna, Q. Yu, M. Boisvert, P. Tijseen and M. Agbandje-McKenna designed research; Q. Yu constructed the pFastBac VP3 clone and prepared the Baculovirus with Bac-to-Bac system. P. Chipman, K.V. Vliet, R. Mckenna and M. Agbandje-McKenna analyzed virus structure data. A manuscript is being prepared for publication.

N.B. References are listed in the Reference section of the main body of the thesis (P. 154)

Abstract

Serpine AAV (sAAV) is the only known parvovirus to infect reptiles. While diverse in sequence, this virus has been grouped into the *Dependovirus* genus because it is most similar to the vertebrate AAVs and parvoviruses of waterfowl. To investigate the capsid features of the sAAV, its structure was determined by cryo-electron microscopy and image reconstruction. The structure built by pseudo-atomic modeling showed that most significant differences were found in loops formed by VR-IV, VP-VIII, and VR-IX. These regions are reported to be determinants of receptor attachment, transduction determinants, and antigenic regions, and identify sAAV as a potential platform for developing AAV vectors which can be altered for specific cell targeting and host antibody immune evasion.

Résumé

Le virus de serpent « Serpine AAV » (sAAV) est le seul parvovirus de reptile connu. Malgré sa divergence de séquence, ce virus a été classé dans le genre *Dependovirus* puisqu'il est similaire au virus AAV de sauvagine. Pour investiguer les caractéristiques de la capsid du virus sAAV, sa structure a été déterminée par cryo-microscopie électronique et reconstruction d'image. La structure a été obtenue par modélisation pseudo-atomique et a montré que les différences les plus significatives étaient dans les régions variables VR-IV, VR-VIII, VR-IX. Ces régions ont été démontrées comme étant des déterminants pour l'attachement au récepteur, pour la transduction et comme étant des régions antigéniques. Ceci a aussi identifié sAAV comme une plateforme potentielle pour développer des vecteurs AAV qui peuvent être altérés pour ciblée des types cellulaires spécifiques, et pour éviter la réponse immunitaire.

Results

The isolation and successful propagation of a reptilian parvovirus-like virus in tissue culture has been reported, parvovirus-like particles were often detected in the presence of adenovirus and herpesvirus or other pathogens, such as picornaviruses (Ahne *et al.*, 1989). In 2004, the genome of a parvovirus isolated from royal python (*Python regius*) was cloned and sequenced (Farkas *et al.*, 2004). The full-length genome was 4432-nt-long and flanked by inverted terminal repeats (ITRs) of 154 nt, containing 122 nt terminal hairpins. Two large open reading frames encoded the non-structural and structural proteins. Analysis of the genome indicated this was most similar to those from adeno-associated viruses and should be classified into the genus *Dependovirus* as Serpine adeno-associated virus (sAAV).

The Adeno-associated viruses (AAVs) belong to the *Dependovirus* genus of the *Parvoviridae* family and are considered to be replication deficient due to a requirement for a helper virus, such as adenovirus or herpesvirus, for genome expression and replication. These viruses have become important therapeutic gene delivery vectors in recent years (Drouin *et al.*, 2013). Currently, 13 distinct human and nonhuman primate AAV serotypes (AAV1–AAV13) have been sequenced, PCR studies of both nonhuman primate and human tissues have identified numerous other AAV genomes (Figure a1.9 and Table a1.6).

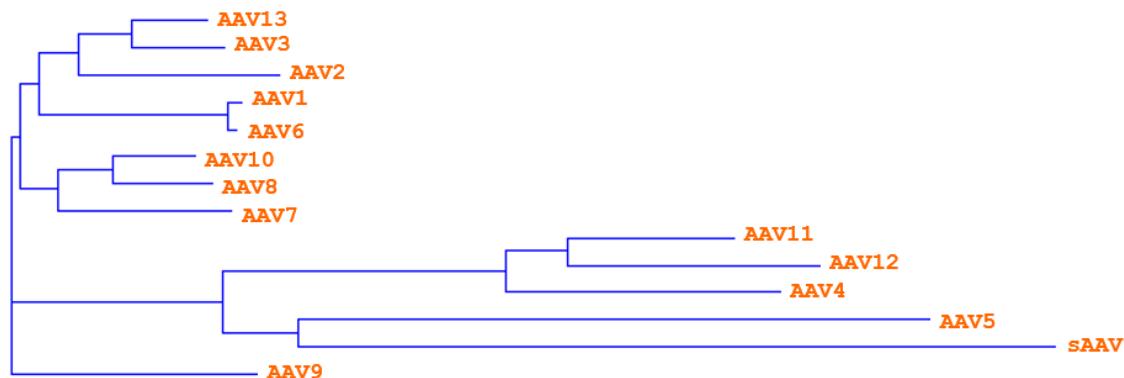


Figure a1.9 Phylogenetic Tree for AAV1-AAV13 and sAAV.

Table a1.6 AAV sequence identities.

	AAV2	AAV4	AAV5	AAV8	AAV9	sAAV
AAV2	100	58	58	82	81	56
AAV4		100	53	59	57	53
AAV5			100	58	57	53
AAV8				100	83	57
AAV9					100	57

The AAV serotypes demonstrate unique tissue tropisms, as well as other characteristics including differential transduction efficiencies and blood-clearance properties (Asokan *et al.*, 2010, Asokan *et al.*, 2012, Kotchey *et al.*, 2011, Mingozi *et al.*, 2011). These unique properties are likely dictated by their viral capsid structure. Using x-ray crystallography or cryo-electron microscopy and image reconstruction (cryo-reconstruction), the structures of AAV1–AAV9 have been determined (DiMattia *et al.*, 2012, Govindasamy *et al.*, 2013, Govindasamy *et al.*, 2006, Kronenberg *et al.*, 2001, Lerch *et al.*, 2010, Nam *et al.*, 2011, Nam *et al.*, 2007, Ng *et al.*, 2010, Walters *et al.*, 2004, Xie *et al.*, 2002). Here we report the structure of sAAV determined by cryo-electron microscopy and image reconstruction. The medium resolution of the reconstructed image, 6.7 Å, reveals a capsid topology that conserves the features reported for other AAVs, including depressions at the 2-fold and surrounding the icosahedral 5-fold axes and three separate protrusions surrounding the 3-fold axes. A cylindrical channel at the icosahedral 5-fold axis connects the interior and exterior surfaces of the capsid (Figure a1.10).

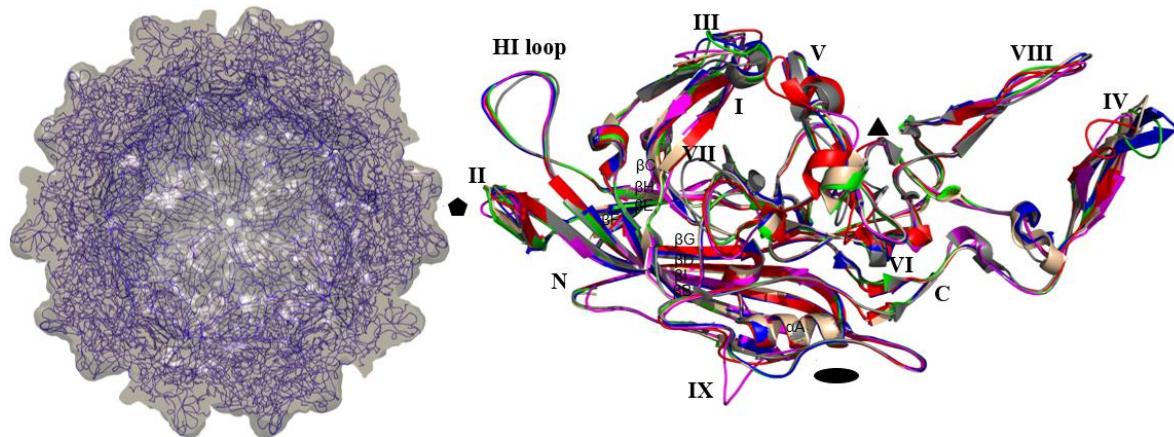


Figure a1.10 Pseudo-atomic model of sAAV.

Left: Coil representation of sAAV VP3 model (in purple) docked into reconstructed density (in semi-transparent grey). View is approximately down the 5-fold axis and shows fit of the β BIDG sheet inside the capsid. Right: Superposition of the cartoon representation of sAAV (purple) onto AAV2 (blue), AAV4 (red), AAV5 (grey), AAV8 (green) and AAV9 (brown). Eight-stranded β -barrel (β B-I), α A, and variable regions (I-IX) are labeled.

As previously reported for other the vertebrate AAVs, only the VP3 common region of the overlapping VP1, 2, and 3 capsid viral protein chains is ordered in the structure and could be built by pseudo-atomic modeling (using the AAV5 crystal structure) (Venkatakrishnan *et al.*, 2013). A comparison to several available AAV structures: AAV2, AAV4, AAV5, AAV8, and AAV9, showed the most significant difference at the 2-fold region due to an insertion and conformational change in the previously defined variable region (VR) IX which forms the wall of the depression. VR-IV (tip of 3-fold protrusions) is short as observed in AAV5. Minor variations are observed in VR-II (top of 5-fold cylinder), and VR-VIII (inner wall of 3-fold protrusions) (Figure a1.10 and a1.11).

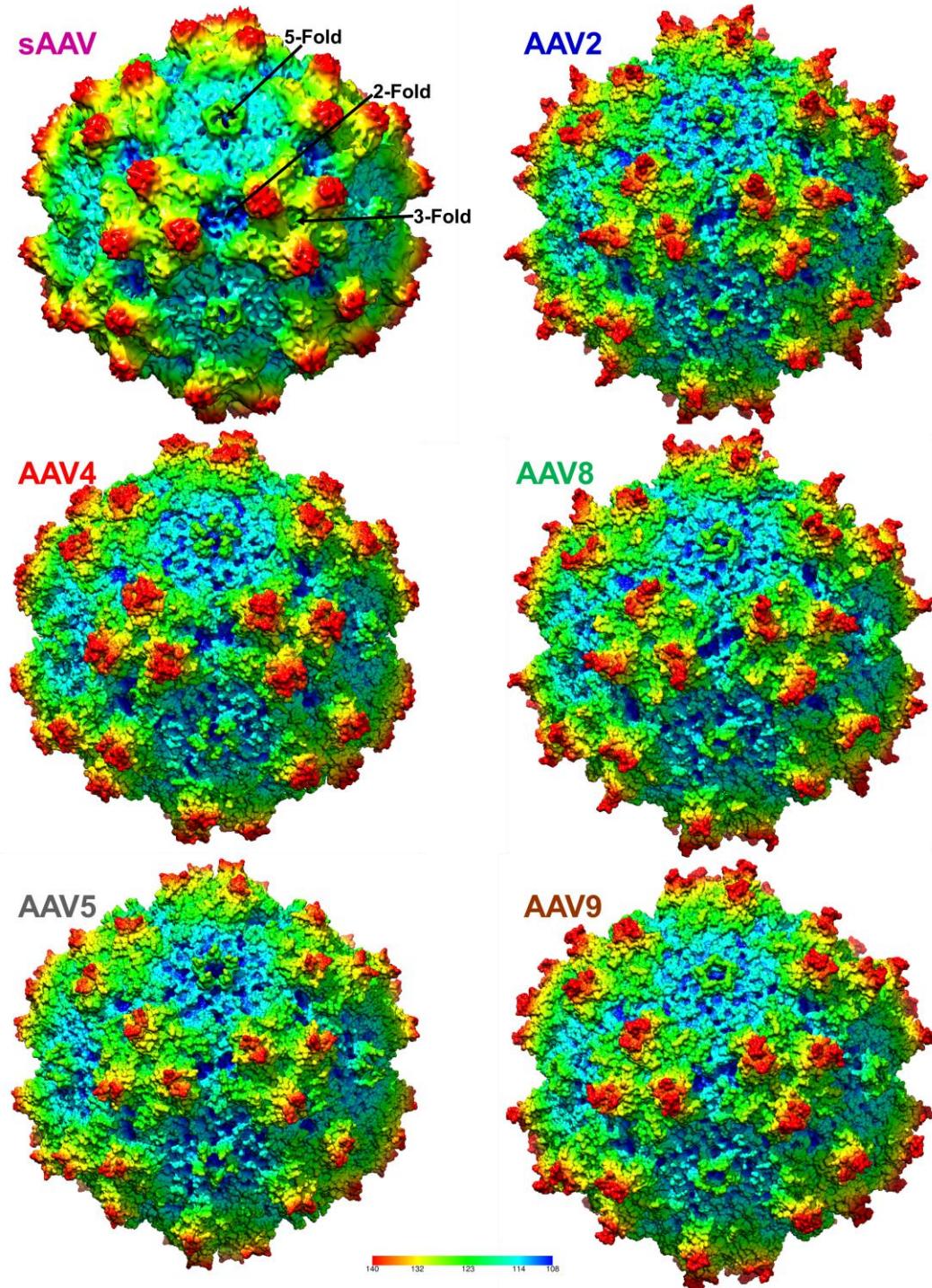


Figure a1.11 AAV capsid structures.

A conformational difference at VR-IX results in a wider 2-fold depression for sAAV. sAAV has less pronounced and rounder 3-fold protrusions due to change in VR-IV, grouping it with AAV4, AAV5, and AAV9 with respect to this surface feature. A VR-II difference alters the top of the 5-fold channel. Color key in Å.

Appendix II: Novel single-stranded DNA viruses

Article 9: Organization of the Ambisense Genome of the *Helicoverpa armigera* Densovirus

Mohamed El-Far,* Jozsef Szelei, Qian Yu, Gilles Fédière,* Max Bergoin, and Peter Tijssen

INRS-Institut Armand-Frappier, Laval, Quebec, Canada

Contribution of authors

Mohamed El-Far was the main investigator as part of his Ph.D study on HaDNV. Jozsef Szelei participated in the cloning steps. Qian Yu participated in the DNA preparation and sequence analysis. P. Tijssen prepared the manuscript. Gilles Fédière and Max Bergoin isolated the virus in Egypt. P. Tijssen supervised the project.

Abstract

A natural densovirus (DNV) of a serious phytophagous pest, *Helicoverpa armigera*, was isolated. The genome of HaDNV contained 6,039 nucleotides (nt) and included inverted terminal repeats (ITRs) of 545 nt with terminal Y-shaped hairpins of 126 nt. Its DNA sequence and ambisense organization with four typical open reading frames (ORFs) demonstrated that it belonged to the genus *Densovirus** in the subfamily *Densovirinae* of the family *Parvoviridae*.

Résumé

Un densovirus (DNV) naturel du ravageur majeur polyphage, *Helicoverpa armigera*, a été isolé. Le génome du virus HaDNV contient 6039 nucléotides (nt), incluant les répétitions terminales inversées (ITRs) de 545nt avec des épingle à cheveux en Y de 126nt. La séquence du génome a montré une organisation ambisens avec quatre cadres de lecture ouverts typiques qui appartient au genre *Densovirus** dans la sous famille *Densovirinae* de la famille *Parvoviridae*.

* Note: Since this manuscript was published, the genus *Densovirus* was renamed as *Ambidensovirus* (Cotmore et al., 2014a).

N.B. References are listed in the Reference section of the main body of the thesis (P. 154)

Results

The cotton bollworm, *Helicoverpa armigera* (Hübner) (Noctuidae), is one of the most serious polyphagous pest species, causing huge losses predominantly in the Old World. Economic damage is greatest in cotton and vegetables. In grain legumes, which are staple foods for people in many countries, up to 80% of the crop can be destroyed. The cotton bollworm also attacks a great number of cereal, vegetable, and garden crops.

The cotton bollworm has become increasingly resistant to pesticides (Gunning et al., 2005), necessitating biological or integrated pest management. Densoviruses (DNVs) have shown great potential, e.g., complete control of *Casphalia extranea* in the oil palm industry in Ivory Coast (Fediere, 1996). Since the late 1980s, we have been developing a program to obtain natural DNV isolates from pests in agricultural fields in Egypt. This program led previously to the isolation of the MIDNV (a pathogen of the maize worm) (Fédière et al., 2004). Here we report the isolation, cloning, and sequence analysis of a densovirus, HaDNV, from the cotton bollworm. Infected larvae stopped feeding within a few days and died within a week. The virus remained infectious for months in cadavers in the field.

HaDNV DNA, extracted from the virus under conditions of high ionic strength to anneal the single-stranded DNA (ssDNA), has a size of around 6 kb. This DNA was blunt-ended by a mixture of Klenow fragment and T4 DNA polymerase and cloned into the *EcoRV* site of the pBluescript vector. The restriction profiles of the viral DNA indicated a close relationship to both MIDNV and *Galleria mellonella* densovirus

(GmDNV). Digesting viral DNA with *Bam*HI gave rise to a 5.5-kb fragment and a 0.3-kb fragment. This small DNA fragment was probably a doublet, resulting from the existence of symmetrical *Bam*HI sites within the viral putative inverted terminal repeats (ITRs), similar to fragments from some other densoviruses (Fédière *et al.*, 2004, Tijssen *et al.*, 2011, Tijssen *et al.*, 2003). This observation helped to clone the viral DNA by methods that were successfully employed for GmDNV and MIDNV (Fédière *et al.*, 2004, Tijssen *et al.*, 2003). Two complete clones, pHa2 and pHa8, were sequenced, both in both directions, using Sanger's method and the primer-walking method as described before (Tijssen *et al.*, 2003). The sequences of the two clones were identical.

The HaDNV genome contained inverted terminal repeats (ITRs), typical of members of the *Densovirus* genus, with a length of 545 nucleotides (nt). The terminal Y-shaped hairpins of 126 nt are completely conserved between HaDNV and MIDNV. The overall sequence is 90% identical to the corresponding sequence of MIDNV and 86% identical to that of GmDNV.

The large open reading frame (ORF) 1 (nt 657 to 1352) on the plus strand had a coding capacity for NS1 of 547 amino acids (aa). ORF2 corresponded to NS2 (nt 1359 to 2996) with 275 aa, and ORF3 (nt 1366 to 2193) corresponded to NS3 with 231 aa. On the complementary minus strand, a large ORF (also on the 5= half at nt 3028 to 5463) with a potential coding region of 811 aa corresponded well to those of the VP structural proteins of related densoviruses. The distribution of the putative coding sequences of HaDNV on the two viral strands thus implied an ambisense organization, and HaDNV contained the typical NS-1 helicase superfamily III and VP phospholipase A2 (Zádori *et al.*, 2001) motifs observed in parvoviruses.

Nucleotide sequence accession number. The GenBank accession number of HaDNV is JQ894784.

Article 10: *Pseudoplusia includens* Densovirus Genome Organization and Expression Strategy

Oanh T. H. Huynh, Hanh T. Pham, Qian Yu, and Peter Tijssen

INRS-Institut Armand-Frappier, Laval, Quebec, Canada

Contribution of authors

Oanh Huynh was the main investigator as part of her M.Sc. study on PiDNV. Since there were many difficulties in the obtaining of hairpin sequences of these viruses, Hanh Pham developed an alternative method to amplify and clone the terminal sequences and Qian Yu participated in the cloning steps. P. Tijssen supervised the project.

Abstract

The genome of a densovirus of a major phytophagous pest, *Pseudoplusia includens*, was analyzed. It contained 5,990 nucleotides (nt) and included inverted terminal repeats of 540 nt with terminal Y-shaped hairpins of 120 nt. Its DNA sequence and ambisense organization with 4 typical open reading frames demonstrated that it belonged to the genus *Densovirus** in the subfamily *Densovirinae* of the family *Parvoviridae*.

Résumé

Le génome du densovirus du ravageur majeur de soja, *Pseudoplusia includens*, a été analysé. Il contient 5990 nucléotides (nt), incluant les répétitions terminales inversées (ITRs) de 540nt avec des épingle à cheveux en Y de 120nt. La séquence du génome a montré une organisation ambisens avec quatre cadres de lecture ouverts typiques qui appartient au genre *Densovirus** dans la sous famille *Densovirinae* de la famille *Parvoviridae*.

* Note: Since this manuscript was published, the genus *Densovirus* was renamed as *Ambidensovirus* (Cotmore et al., 2014a).

N.B. References are listed in the Reference section of the main body of the thesis (P. 154)

Results

The distribution of the polyphagous soybean looper pest, *Pseudoplusia includens* (syn., *Chrysodeixis includens* [Hübner] [Noctuidae, Plusiinae, Lepidoptera]), is restricted to the Western Hemisphere, occurring from southern Canada to southern South America (Alford et al., 1982). In addition to the soybean, it may feed on a large number of crops of economic importance (Macrae et al., 2005, McPherson et al., 2009). Previously, two small icosahedral viruses have been isolated from the soybean looper, a picornavirus and a smaller virus with biophysical properties that seem to match those of the densoviruses (Chao et al., 1985).

Densoviruses are notoriously unstable upon cloning (Liu et al., 2011, Tijssen et al., 2006a, Tijssen et al., 2003, Yu et al., 2012a, Yu et al., 2012b), and densovirus entries in GenBank, such as those from *Junonia coenia* (JcDNV) (Dumas et al., 1992) and *Diatraea saccharalis* (DsDNV) (NC_001899), often lack significant parts of their inverted terminal repeats (ITRs). DNA purified from *Pseudoplusia includens* DNV (PiDNV) in phosphate-buffered saline (PBS) had a size of around 6 kb. This DNA was blunt ended by a mixture of Klenow fragment and T4 DNA polymerase and cloned into a linear pJazz vector (from Lucigen Corp.), which lacks transcription into the insert and torsional stress (Godiska et al., 2009) to prevent recombination and deletion of insert fragments. Six clones, or about 0.3%, had full-length inserts and could be stably subcloned into circular vectors.

Four complete clones were sequenced in both directions, using Sanger's method and the primer-walking method as described before (Tijssen et al., 2003), and

the contigs were assembled by the CAP3 program (<http://pbil.univ-lyon1.fr/cap3.php/>) (Huang *et al.*, 1999). The difficulties encountered with sequencing of the terminal hairpins were solved by sequencing after (i) digestion near the middle of the hairpin with *Bst*UI restriction enzyme or (ii) amplifying the hairpins by PCR in the presence three additives: 1.3 M betaine, 5% dimethyl sulfoxide, and 50 mM 7-deaza-dGTP. Sequences of the clones, except for the flip-flop regions in the hairpins, were identical. In the hairpins, nucleotides (nt) 46 to 75 and nt 5916 to 5945 occurred in two orientations, “flip” and its reverse complement orientation “flop.” The ambisense PiDNV genome contained typical ITRs of members of the *Densovirus* genus with a length of 540nt and terminal Y-shaped hairpins of 120 nt. The overall sequence of 5,990 nt was 83 to 87% identical with those of other viruses in the *Densovirus** genus but about 50 nt shorter.

The open reading frames (ORFs) were conserved with members of the *Densovirus** genus, and the putative splicing sites were conserved with those that have been identified for *Galleria mellonella* DNV (GmDNV) (Tijssen *et al.*, 2003) and *Mythimna loreyi* DNV (MIDNV) (Fédière *et al.*, 2004). The large ORF1 (nt 1355 to 3019) on the plus strand had a coding capacity for NS1 of 554 amino acids (aa), ORF2 corresponded to NS2 (nt 1362 to 3019) with 275 aa, and ORF3 (nt 647 to 1348) corresponded to NS3 with 233 aa. On the complementary minus strand, a large ORF (also on the 5= half at nt 3006 to 5423) with a potential coding region of 805 aa corresponded well to those of the VP structural proteins of related densoviruses. The distribution of the putative coding sequences implied an ambisense organization and expression, and PiDNV contained the typical NS-1 helicase superfamily III and VP phospholipase A2 (Zádori *et al.*, 2001) motifs observed in other parvoviruses.

Nucleotide sequence accession number. The GenBank accession number of PiDNV is JX645046.

ACKNOWLEDGMENTS

This work was supported by the Natural Sciences and Engineering Research Council of Canada to P.T. O.T.H.H. acknowledges support from a scholarship from the Biochemical Institute of Ho Chi Minh City, Vietnam. Q.Y. acknowledges support from a scholarship from the People's Republic of China. O.T.H.H., H.T.P., and Q.Y. acknowledge tuition waivers at INRS.

Article 11: A Circo-Like Virus Isolated from *Penaeus monodon* Shrimps

Hanh T. Pham,^a Qian Yu,^a Maude Boisvert,^a Hanh T. Van,^b Max Bergoin,^a Peter Tijssen^a

INRS, Institut Armand-Frappier, Laval, QC, Canada ^a; Institute of Tropical Biology, Vietnam Academy of Science and Technology, Ho Chi Minh City, Vietnam ^b

Contribution of authors

Hanh Pham was mainly responsible for this work including virus purification from infected crickets, DNA extraction, cloning, and result analysis. Qian Yu and Maude Boisvert have contributed in the cloning steps, sequences and result analysis. Hanh Thi Van (Vietnam) sent us shrimps that contained a circo-like virus. Peter Tijssen and Max Bergoin planned and supervised the project. Hanh Pham and Peter Tijssen prepared the manuscripts.

N.B. References are listed in the Reference section of the main body of the thesis (P. 154)

Abstract

A virus with a circular Rep-encoding single-stranded DNA (ssDNA) (CRESS-DNA) genome (PmCV-1) was isolated from *Penaeus monodon* shrimps in Vietnam. The gene structure of the 1,777-nucleotide (nt) genome was similar to that of circoviruses and cycloviruses, but the nucleic acid and protein sequence identities to these viruses were very low.

Résumé

Un virus (PmCV-1) avec un génome circulaire d'ADN simple brin (CRESS-DNA) encodant les protéines Rep a été isolé à partir de crevettes *Penaeus monodon* du Vietnam. La structure des gènes du génome de 1777 nucléotides (nt) était similaire aux circovirus et cyclovirus mais les identités de séquences nucléotidiques et protéiques étaient très faibles.

Results

Recently, viral metagenomics revealed circo-like viruses in the marine copepod species *Acartia tonsa* and *Labidocera aestival* (Crustaceae) (Dunlap *et al.*, 2013). Here, we report the isolation by classical methods of a similar virus with a circular Rep-encoding single-stranded DNA (ssDNA) (CRESS-DNA) genome from *Penaeus monodon* shrimps (PmCV-1).

Circoviruses are nonenveloped, icosahedral particles and contain circular ssDNA genomes of about 1.7 to 2.3 kb. The open reading frame (ORF) for the Rep protein codes for conserved rolling circle replication (RCR) and superfamily 3 (SF3) helicase motifs (Delwart *et al.*, 2012, Rosario *et al.*, 2012). In contrast, the cap gene is generally not conserved. Originally, circoviruses were isolated from pig and bird species (Ritchie *et al.*, 1989, Tischer *et al.*, 1986, Todd *et al.*, 1991), but in vitro rolling circle amplification, high-throughput sequencing, and metagenomics studies have led to rapid expansion of the known diversity and host range of small CRESS-DNA viruses (CVs). This also led to an unsettled viral taxonomy with different subfamilies within the *Circoviridae* family and reassignment of their members (Rosario *et al.*, 2012).

In this study, about 100 g of cleaned, diseased *Penaeus monodon* shrimps was homogenized and virus was purified (isolate VN11 from Vietnam) as described previously (Pham *et al.*, 2013c, Pham *et al.*, 2013e). Viral DNA was isolated from purified viruses with the High Pure viral nucleic acid kit (Roche Applied Science), followed by rolling circle amplification by ϕ 29 DNA polymerase (NEB) at 30°C for 6 hours (de Vega *et al.*, 2010). Amplified product was then digested with EcoRI and separated on a 1% agarose gel. A band of 1.8 kb was recovered from the gel and cloned into a pBluescript-KS(+) vector. Clones were sequenced by Sanger's method and primer walking. PCR with outward primers was carried out and the amplicon was cloned into a TA vector (pGEMT-easy, Promega). All sequencing results were assembled using the CAP3 program (Huang *et al.*, 1999).

Sequence analysis revealed that PmCV-1 is closely related to members of the *Cyclovirus* genus in the *Circoviridae* family. PmCV-1 possesses a 1,777-nucleotide (nt) genome containing three ORFs encoding 266, 255, and 146 amino acids (aa). Numbering starts with the loop in the conserved stem loop. The 266-aa product of the largest ORF, from nt 51 to 851, shared about 30% sequence identity (over 90% of query coverage) with the putative Rep of cycloviruses and contained RCR and SF3 motifs. The 255-aa product of the ORF translated in the opposite direction, from nt 1,671 to 904, shared 25% identity with the Cap protein of a *Diporeia* sp.-associated circular-virus (GenBank accession no. KC248415.1, E value 0.004), and thus the ORF probably encodes the capsid protein. The smallest ORF, from nt 1,246 to 1,686, codes for a 146-aa protein that did not reveal any amino acid similarity using Blastx in a protein database with E values of < 0.01. The 156-nt intergenic region between the 5' ends of putative cap and rep genes encompasses 13-nt inverted repeats (nt 11 to 23 and 1765 to 1777) forming a stem and a 10-nt loop containing a canonical nonanucleotide, TAATATTAC, between nt 2 and 10. The intergenic region between the 3' ends of the cap and rep genes is 53 nt long. The genome structure resembles that of circoviruses and cycloviruses.

Metagenomic discovery has particularly impacted the discovery of CRESS-DNA viruses, both in host range and genetic diversity. Although this approach is very powerful, its perils should not be underestimated (Naccache *et al.*, 2013).

Nucleotide sequence accession number. The GenBank accession number for PmCV-1 is KF481961.

ACKNOWLEDGMENTS

This work was supported by the Natural Sciences and Engineering Research Council of Canada (NSERC) to P.T.; H.T.P. and Q.Y. acknowledge tuition fee waivers at INRS and scholarships from INRS and the People's Republic of China. M.B. acknowledges a scholarship from NSERC.

Appendix III: Publications

Iteravirus-Like Genome Organization of a Densovirus from *Sibine fusca* Stoll

Qian Yu,^a Gilles Fédière,^{a,*} Adly Abd-Alla,^b Max Bergoin,^a and Peter Tijssen^a

INRS-Institut Armand-Frappier, Laval, Quebec, Canada,^a and Insect Pest Control Laboratory, Joint FAO/IAEA Division, International Atomic Energy Agency, Vienna, Austria^b

The complete genome of *Sibine fusca* densovirus was cloned and sequenced. The genome contained 5,012 nucleotides (nt), including inverted terminal repeats (ITRs) of 230 nt with terminal hairpins of 161 nt. Its DNA sequence and monosense organization with 3 open reading frames (ORFs) is typical of the genus *Iteravirus* in the subfamily *Densovirinae* of the *Parvoviridae*.

The slug caterpillar *Sibine fusca* Stoll (syn. *Acharia fusca*; Limacodidae), a major pest of oil palm, is widely distributed in Southern America. The larvae live gregariously in colonies of up to 60 individuals and disperse just before pupation. Virus-infected larvae showed nuclear lesions typical of a densovirus infection (2, 8). Electron microscopy revealed isometric particles with about 20-nm diameters, also characteristic of densoviruses. Infected larvae of *S. fusca* drifted from the feeding colony, and considerable proliferation of cells, resembling tumors, was observed in the midgut (8). Thus far, this virus has not been further characterized, cloned, or sequenced but has been used effectively in biological control (4).

The virus was partially purified by the method described for *Galleria mellonella* densovirus (GmDNV) (7) from an infected larva. A sequence-independent, single-primer amplification (SISPA) method (9) was used in a preliminary genome characterization. DNA, extracted under conditions of high ionic strength to anneal the single-stranded DNA (ssDNA), had a size of around 5 kb. This DNA was digested with the Csp6I restriction enzyme, ligated with an adaptor, amplified by PCR as described elsewhere (1), and cloned into the PCR2.1 vector by the TA cloning method (5). Amplicon inserts were sequenced by Sanger's method as described previously (11). A unique ClaI restriction site was observed near the middle of a preliminary 4.7-kb sequence. DNA from the virus was then blunt-ended by a mixture of Klenow fragment and T4 DNA polymerase, digested with ClaI, and cloned into EcoRV and ClaI sites in the pBluescriptSK(–) vector, yielding clones with a 2.6-kb insert and clones with a 2.4-kb insert. Four inserts of each set were sequenced in both directions using Sanger's method and the primer-walking method as described before (11). Insert sequences were identical in each set except for the flip-flop sequences in the hairpins.

The *Sibine fusca* densovirus (SfDNV) genome contained inverted terminal repeats (ITRs) typical of the three members (*Bombyx mori* densovirus type 1 [BmDNV-1], *Caspalia extranea* densovirus [CeDNV], and *Dendrolimus punctatus* densovirus [DpDNV]) of the *Iteravirus* genus and with a length of 230 nucleotides (nt) (10). The terminal J-shaped hairpins of 161 nt were about 90% conserved between BmDNV-1 (6), CeDNV (3), and DpDNV (12). In the hairpins, nt 60 to 102 and nt 4911 to 4953 occurred in two orientations, “flip” and its reverse complement orientation “flop,” that were identical to the flip-flop of CeDNV and 98% identical to that of BmDNV. The overall sequence was about 85% identical to CeDNV, about 78% identical to BmDNV, and about 72% identical to DpDNV.

The monosense genome contained three intronless genes that were virtually identical in position and size to those of other iteraviruses. The largest open reading frame (ORF), ORF1 (nt 354 to 2615), had a coding capacity of 753 amino acids (aa) and the typical NTPase motif for NS1 (3). ORF2 (nt 2669 to 4714), with the phospholipase A2 motif characteristic for VP (13), had a coding capacity of 681 aa. ORF3 corresponded to NS2 with a 452-aa coding capacity and typically overlapped the N terminus of NS1 (nt 481 to 1839). As a comparison, for the other iteraviruses, the NS1 is 753 to 775 aa, the NS2 is 451 to 453 aa, and the VP is 668 to 678 aa.

Nucleotide sequence accession number. The GenBank accession number of SfDNV is JX020762.

ACKNOWLEDGMENTS

This work was supported by a grant from the Natural Sciences and Engineering Research Council of Canada to P.T. Q.Y. acknowledges support from a scholarship from the People's Republic of China. G.F. was supported by IRD during his sabbatical at INRS. A. A.-A. is supported by IAEA.

G.F. and M.B. are invited professors at INRS.

REFERENCES

- Allander T, Emerson SU, Engle RE, Purcell RH, Bukh J. 2001. A virus discovery method incorporating DNase treatment and its application to the identification of two bovine parvovirus species. Proc. Natl. Acad. Sci. U. S. A. 98:11609–11614.
- Fediére G. 2000. Epidemiology and pathology of Densovirinae, p 1–11. In Faisst S, Rommelaeire J (ed), Parvoviruses. Karger, Basel, Switzerland.
- Fediére G, Li Y, Zadori Z, Szelei J, Tijssen P. 2002. Genome organization of *Caspalia extranea* densovirus, a new iteravirus. Virology 292:299–308.
- Genty P, Mariaud D. 1975. Utilisation d'un germe entomopathogène dans la lutte contre *Sibine fusca* (Limacodidae). Oleagineux 30:349–354.
- Holton TA, Graham MW. 1991. A simple and efficient method for direct cloning of PCR products using ddT-tailed vectors. Nucleic Acids Res. 19:1156.
- Li Y, et al. 2001. Genome organization of the densovirus from *Bombyx mori* (BmDNV-1) and enzyme activity of its capsid. J. Gen. Virol. 82: 2821–2825.
- Longworth JF, Tinsley TW, Barwise AH, Walker IO. 1968. Purification of a non-occluded virus from *Galleria mellonella*. J. Gen. Virol. 3:167–174.

Received 20 May 2012 Accepted 22 May 2012

Address correspondence to Peter Tijssen, peter.tijssen@iaf.inrs.ca.

* Present address: Gilles Fédière, IRD, Nouméa, New Caledonia.

Copyright © 2012, American Society for Microbiology. All Rights Reserved.

doi:10.1128/JVI.01267-12

8. Meynadier G, Amargier A, Genty P. 1977. Une virose de type denso-nucléose chez le lépidoptère *Sibine fusca* Stoll. *Oleagineux* 32:357–361.
9. Reyes GR, Kim JP. 1991. Sequence-independent, single-primer amplification (SISPA) of complex DNA populations. *Mol. Cell. Probes* 5:473–481.
10. Tijssen P, et al. 2011. Parvoviridae, p 375–395. In King AMQ, Adams MJ, Carstens E, Lefkowitz EJ (ed), *Virus taxonomy: classification and nomenclature of viruses: ninth report of the International Committee on Taxonomy of Viruses*. Elsevier, San Diego, CA.
11. Tijssen P, et al. 2003. Organization and expression strategy of the ambisense genome of denso-nucleosis virus of *Galleria mellonella*. *J. Virol.* 77:10357–10365.
12. Wang J, et al. 2005. Nucleotide sequence and genomic organization of a newly isolated densovirus infecting *Dendrolimus punctatus*. *J. Gen. Virol.* 86:2169–2173.
13. Zadori Z, et al. 2001. A viral phospholipase A2 is required for parvovirus infectivity. *Dev. Cell* 1:291–302.

Papilio polyxenes Densovirus Has an Iteravirus-Like Genome Organization

Qian Yu,^a Ann E. Hajek,^b Max Bergoin,^a and Peter Tijssen^a

INRS-Institut Armand-Frappier, Laval, Quebec, Canada,^a and Cornell University, Department of Entomology, Ithaca, New York, USA^b

The genome of *Papilio polyxenes* densovirus was cloned and sequenced and contained 5,053 nucleotides (nt), including inverted terminal repeats (ITRs) of 271 nt with terminal hairpins of 175 nt. Its DNA sequence and monosense organization with 3 open reading frames (ORFs) are typical of the genus *Iteravirus* in the subfamily *Densovirinae* of the *Parvoviridae*.

The larvae of the black swallowtail (*Papilio polyxenes*; family *Papilionidae*), a butterfly found throughout eastern North America, feed gregariously on host plants of the carrot family (*Umbelliferae*), such as dill, parsley, and fennel. These plants produce stimulatory compounds for chemoreception of this insect but also furanocoumarins that are toxic and serve as a defense mechanism against various insect predators. *Papilio polyxenes*, which adapted to furanocoumarin-containing host plants, therefore provides an excellent laboratory model to study insect-plant coevolution (1, 10). During recent years, significant mortality was observed in laboratory animals and electron microscopy examination revealed isometric particles of about 20 nm in diameter (F. Pringle, unpublished observations), characteristic of densovirus. A significant proportion of the larvae obtained from the field were also infected.

Papilio polyxenes densovirus (PpDNV) was partially purified from 0.5 g larvae as described previously for *Galleria mellonella* DNV (GmDNV) (5) and *Sibine fusca* Stoll DNV (SfDNV) (11). PpDNV DNA, extracted under conditions of high ionic strength to anneal the single-stranded DNA (ssDNA), had a size of around 5 kb. A sequence-independent, single-primer amplification (SISPA) method (6) was used as previously described (11). The amplicons were cloned into the pGEM-T vector by the TA cloning method (3) and sequenced by Sanger's method as described previously (8). A unique SacI restriction site detected near the middle of a 4.7-kb sequence was used to clone the blunt-ended PpDNV DNA, obtained from virus, restricted with SacI in the EcoRV and SacI sites of the PCR2.1 vector. Clones with a 1.6-kb insert and clones with a 3.4-kb insert were obtained, and four inserts of each set were sequenced in both directions using Sanger's method and the primer-walking method as described previously (8). Insert sequences were identical in each set except for the flip-flop sequences.

The overall sequence had a high identity with iteraviruses (identity of about 78% to *Casphalia extranea* DNV [CeDNV], about 75% to *Bombyx mori* DNV [BmDNV-1], about 74% to SfDNV, and about 67% to *Dendrolimus punctatus* DNV [DpDNV]). The PpDNV genome contained typical inverted terminal repeats (ITRs) of the four members of the *Iteravirus* genus (BmDNV-1, CeDNV, SfDNV, and DpDNV), albeit a bit longer (271 versus 230 nucleotides [nt]) (7). The terminal J-shaped hairpins of 175 nt were about 80% conserved with BmDNV-1 (4), CeDNV (2), SfDNV (11), and DpDNV (9). In the hairpins, nt 67 to 109 and nt 4945 to 4987 occurred in two orientations, "flip" and its reverse-complement orientation "flop," that were close to 100% identical to the flip-flop of the other iteraviruses.

Similar to other iteraviruses, the PpDNV monosense genome contained three intronless genes with essentially identical positions and sizes. The largest, open reading frame 1 (ORF1) (nt 349 to 2631), had a coding capacity of 760 amino acids (aa) and the typical nucleoside triphosphatase (NTPase) motif for NS1 (2). ORF2 (nt 2686 to 4707) with the phospholipase A2 motif typical for parvovirus VP (12) had a coding capacity of 673 aa. ORF3 corresponded to NS2 with a 455-aa coding capacity and typically overlapped the N terminus of NS1 (nt 482 to 1849). As a comparison, NS1 is 753 to 775 aa, NS2 is 451 to 453 aa, and VP is 668 to 678 aa for the other iteraviruses.

Nucleotide sequence accession number. The GenBank accession number for PpDNV is JX110122.

ACKNOWLEDGMENTS

This work was supported by a grant from the Natural Sciences and Engineering Research Council of Canada to P.T. Q.Y. acknowledges support from a scholarship from the People's Republic of China. M.B. is an invited professor at INRS.

REFERENCES

1. Berenbaum M, Feeny P. 1981. Toxicity of angular furanocoumarins to swallowtail butterflies: escalation in a coevolutionary arms race? *Science* 212:927–929.
2. Fédière G, Li Y, Zadori Z, Szeli J, Tijssen P. 2002. Genome organization of *Casphalia extranea* densovirus, a new iteravirus. *Virology* 292:299–308.
3. Holton TA, Graham MW. 1991. A simple and efficient method for direct cloning of PCR products using ddT-tailed vectors. *Nucleic Acids Res.* 19:1156.
4. Li Y, et al. 2001. Genome organization of the densovirus from *Bombyx mori* (BmDNV-1) and enzyme activity of its capsid. *J. Gen. Virol.* 82: 2821–2825.
5. Longworth JF, Tinsley TW, Barwise AH, Walker IO. 1968. Purification of a non-occluded virus from *Galleria mellonella*. *J. Gen. Virol.* 3:167–174.
6. Reyes GR, Kim JP. 1991. Sequence-independent, single-primer amplification (SISPA) of complex DNA populations. *Mol. Cell. Probes* 5:473–481.
7. Tijssen P, et al. 2011. *Parvoviridae*, p 375–395. In King AMQ, Adams MJ, Carstens E, Lefkowitz EJ (ed), *Virus taxonomy: classification and nomenclature of viruses: ninth report of the International Committee on Taxonomy of Viruses*. Elsevier, San Diego, CA.

Received 1 June 2012 Accepted 11 June 2012

Address correspondence to Peter Tijssen, peter.tijssen@iaf.inrs.ca.

Copyright © 2012, American Society for Microbiology. All Rights Reserved.

doi:10.1128/JVI.01368-12

8. Tijssen P, et al. 2003. Organization and expression strategy of the ambisense genome of densonucleosis virus of *Galleria mellonella*. *J. Virol.* **77**:10357–10365.
9. Wang J, et al. 2005. Nucleotide sequence and genomic organization of a newly isolated densovirus infecting *Dendrolimus punctatus*. *J. Gen. Virol.* **86**:2169–2173.
10. Wen Z, Rupasinghe S, Niu G, Berenbaum MR, Schuler MA. 2006. CYP6B1 and CYP6B3 of the black swallowtail (*Papilio polyxenes*): adaptive evolution through subfunctionalization. *Mol. Biol. Evol.* **23**:2434–2443.
11. Yu Q, Fédière G, Abd-Alla A, Bergoin M, Tijssen P. 2012. Iteravirus-like genome organization of a densovirus from *Sibine fusca* Stoll (SfDNV). *J. Virol.* **86**:8897–8898.
12. Zadòri Z, et al. 2001. A viral phospholipase A2 is required for parvovirus infectivity. *Dev. Cell* **1**:291–302.

Iteradenovirus from the Monarch Butterfly, *Danaus plexippus plexippus*

Qian Yu, Peter Tijssen

INRS, Institut Armand-Frappier, Université du Québec, Laval, Quebec, Canada

The 5,006-nucleotide (nt)-long genome of a new virus from monarch butterfly pupae was cloned and sequenced. It was flanked by inverted terminal repeats (ITRs) of 239 nt with 163-nt hairpins. The monosense genome with three open reading frames is typical of the genus *Iteradenovirus* in the subfamily *Densovirinae* of the family *Parvoviridae*.

Received 25 March 2014 Accepted 3 April 2014 Published 17 April 2014

Citation Yu Q, Tijssen P. 2014. Iteradenovirus from the monarch butterfly, *Danaus plexippus plexippus*. *Genome Announc.* 2(2):e00321-14. doi:10.1128/genomeA.00321-14.

Copyright © 2014 Yu and Tijssen. This is an open-access article distributed under the terms of the Creative Commons Attribution 3.0 Unported license.

Address correspondence to Peter Tijssen, peter.tijssen@iaf.inrs.ca.

Monarch butterflies (*Danaus plexippus plexippus*) migrate from eastern and central North America for overwintering in Mexico. Migration of this emblematic butterfly has been in rapid decline in recent years, prompting the presidents of the United States and Mexico and the Prime Minister of Canada to discuss this problem during a meeting in February 2014. Several factors may be responsible for this trend. The cool and relatively moist high mountain habitats of Oyamel fir forests are ideal for both the firs and the butterflies. The forest canopy and the clustering of the monarchs protect them against freezing (1). Severe logging and climate change threaten these forests, and a massive reforestation effort is under way to reverse this trend. Second, the extensive use of genetically modified herbicide-resistant soybeans and corn may be reducing the number of larval host plants, milkweeds, especially in their main habitat in the Corn Belt (2–4), encouraging the suggestion of a milkweed corridor. However, this has been disputed elsewhere (5). Third, pathogens such as bacteria, parasites, and viruses may affect monarch populations (6–8).

Virus was partially purified from three infected pupae obtained from a butterfly farm in Granby (Quebec, Canada) by the method described for *Galleria mellonella* densovirus (9) and visualized by electron microscopy. A preliminary genome characterization was obtained with the sequence-independent single-primer amplification (SISPA) method (10–12), showing two *Sph*I restriction sites in a preliminary 4.7-kb sequence. Viral DNA was then blunt ended by a mixture of Klenow large-fragment and T4 DNA polymerase, digested with *Sph*I, and cloned into EcoRV and *Sph*I sites in the pBluescriptSK II(-) vector, yielding clones with 3.4-kb inserts and clones with 1.5-kb inserts. Sequences of several complete clones, obtained in both directions with Sanger's method (10, 11), were identical except for the flip-flop sequences in the hairpins. The sequence between the two *Sph*I sites was obtained after PCR amplification with gene-specific primers.

The *D. plexippus plexippus* iteradenovirus (DppIDV) genome contained the typical inverted terminal repeats (ITRs) of members of the *Iteradenovirus* genus (*Bombyx mori* densovirus 1 [BmDNV-1], *Capsalia extranea* densovirus [CeDNV], *Sibine fusca* densovirus [SfDNV], *Papilio polyxenes* densovirus [PpDNV], and *Dendrolimus punctatus* densovirus [DpDNV]) (10, 11, 13–15). The 239-nucleotide (nt) ITRs with 163-nt terminal J-shaped hairpins were about 90% conserved with those of the other iteradenoviruses. The overall sequence was about 86% identical to CeDNV, about 84% identical to SfDNV and BmDNV, about 78% identical to PpDNV, and about 71% identical to DpDNV.

Similar to other iteradenoviruses, the DppIDV monosense genome contained three intronless genes with essentially identical positions and sizes. The largest, open reading frame 1 (ORF1) (nt 360 to 2618), had a coding capacity of 752 amino acids (aa) and the typical nucleoside triphosphatase (NTPase) motif for NS1. ORF2 (nt 2677 to 4710), with the phospholipase A2 motif, typical for parvovirus VP, had a coding capacity of 677 aa. ORF3, with a 451-aa coding capacity (nt 487 to 1842) corresponded to NS2 and overlapped NS1 at its N terminus. As a comparison, NS1 is aa 753 to 775, NS2 is aa 451 to 455, and VP is aa 668 to 681 for the other iteradenoviruses.

Nucleotide sequence accession number. The GenBank accession no. for DppIDV is KF963252.

ACKNOWLEDGMENTS

This work was supported by the Natural Sciences and Engineering Research Council of Canada grant to P.T. Q.Y. acknowledges support from a scholarship from the People's Republic of China and tuition waivers from INRS-IAF.

REFERENCES

- Anderson JB, Brower LP. 1996. Freeze-protection of overwintering monarch butterflies in Mexico: critical role of the forest as a blanket and an umbrella. *Ecol. Entomol.* 21:107–116. <http://dx.doi.org/10.1111/j.1365-2311.1996.tb01177.x>.
- Knight A, Brower LP. 2009. The influence of eastern North American autumnal migrant monarch butterflies (*Danaus plexippus* L.) on continuously breeding resident monarch populations in southern Florida. *J. Chem. Ecol.* 35:816–823. <http://dx.doi.org/10.1007/s10886-009-9655-z>.
- Brower LP, Fink LS, Walford P. 2006. Fueling the fall migration of the monarch butterfly. *Integr. Comp. Biol.* 46:1123–1142. <http://dx.doi.org/10.1093/icb/icl029>.
- Malcolm SB, Cockrell BJ, Brower LP. 1989. Cardenolide fingerprint of

- monarch butterflies reared on common milkweed, *Asclepias syriaca* L. J. Chem. Ecol. 15:819–853.
5. Niiler E. 1999. GM corn poses little threat to monarch. Nat. Biotechnol. 17:1154. <http://dx.doi.org/10.1038/70691>.
 6. Arnott HJ, Smith KM, Fullilove SL. 1968. Ultrastructure of a cytoplasmic polyhedrosis virus affecting the monarch butterfly, *Danaus plexippus*. I. Development of virus and normal polyhedra in the larva. J. Ultrastruct. Res. 24:479–507. [http://dx.doi.org/10.1016/S0022-5320\(68\)80050-4](http://dx.doi.org/10.1016/S0022-5320(68)80050-4).
 7. Bartel RA, Oberhauser KS, De Roode JC, Altizer SM. 2011. Monarch butterfly migration and parasite transmission in eastern North America. Ecology 92:342–351. <http://dx.doi.org/10.1890/10-0489.1>.
 8. de Roode JC, de Castillejo CL, Faits T, Alizon S. 2011. Virulence evolution in response to anti-infection resistance: toxic food plants can select for virulent parasites of monarch butterflies. J. Evol. Biol. 24: 712–722. <http://dx.doi.org/10.1111/j.1420-9101.2010.02213.x>.
 9. Tijssen P, Li Y, El-Far M, Szelei J, Letarte M, Zádori Z. 2003. Organization and expression strategy of the ambisense genome of densonucleosis virus of *Galleria mellonella*. J. Virol. 77:10357–10365. <http://dx.doi.org/10.1128/JVI.77.19.10357-10365.2003>.
 10. Yu Q, Fédière G, Abd-Alla A, Bergoin M, Tijssen P. 2012. Iteravirus-like genome organization of a densovirus from *Sibine fusca* stoll. J. Virol. 86: 8897–8898. <http://dx.doi.org/10.1128/JVI.01267-12>.
 11. Yu Q, Hajek AE, Bergoin M, Tijssen P. 2012. *Papilio polyxenes* densovirus has an iteravirus-like genome organization. J. Virol. 86:9534–9535. <http://dx.doi.org/10.1128/JVI.01368-12>.
 12. Allander T, Emerson SU, Engle RE, Purcell RH, Bukh J. 2001. A virus discovery method incorporating DNase treatment and its application to the identification of two bovine parvovirus species. Proc. Natl. Acad. Sci. U. S. A. 98:11609–11614. <http://dx.doi.org/10.1073/pnas.211424698>.
 13. Fédière G, Li Y, Zádori Z, Szelei J, Tijssen P. 2002. Genome organization of *Capsalia extranea* densovirus, a new iteravirus. Virology 292:299–308. <http://dx.doi.org/10.1006/viro.2001.1257>.
 14. Li Y, Zádori Z, Bando H, Dubuc R, Fédière G, Szelei J, Tijssen P. 2001. Genome organization of the densovirus from *Bombyx mori* (BmDNV-1) and enzyme activity of its capsid. J. Gen. Virol. 82:2821–2825.
 15. Wang J, Zhang J, Jiang H, Liu C, Yi F, Hu Y. 2005. Nucleotide sequence and genomic organization of a newly isolated densovirus infecting *Dendrolimus punctatus*. J. Gen. Virol. 86:2169–2173. <http://dx.doi.org/10.1099/vir.0.80898-0>.

Gene Expression of Five Different Iteradensoviruses: *Bombyx mori* Densovirus, *Casphalia extranea* Densovirus, *Papilio polyxenes* Densovirus, *Sibine fusca* Densovirus, and *Danaus plexippus* Densovirus

Qian Yu, Peter Tijssen

INRS-Institut Armand-Frappier, Université du Québec, Laval, Québec, Canada

Iteradensoviruses are 5-kb parvoviruses with typical J-shaped inverted terminal repeats of about 250 nucleotides and terminal hairpins of about 165 nucleotides. The single-stranded DNA genome contains several open reading frames, but their expression strategy is still unknown. Here the transcription maps and expression of the viruses in this genus were explored. As for brevi-densoviruses, the two nonstructural (NS) genes were expressed by overlapping promoters with alternate transcription starts at both sides of the NS1 start codon.

Invertebrate densoviruses (DVs) form a separate subfamily (*Densovirinae*) within the *Parvoviridae* family (1–4) and have recently been reclassified (1). They share many physicochemical and genome properties with *Parvovirinae* of vertebrates (5–7), and most viruses of these subfamilies have a phospholipase A2 (PLA2) activity for cell entry (8–10). The left-hand side of the genome contains the nonstructural (NS) genes with the rolling-circle replication endonuclease and the superfamily 3 helicase in NS1 (11, 12). Open-reading frames (ORFs) for structural proteins (VPs) are on the right-hand side of the genomes (5, 6).

Densovirinae include the *Ambidensovirus*, *Brevidensovirus*, *Hepandenvirus*, *Penstyldensovirus*, and *Iteradensovirus* genera, which have distinct genome sizes and organizations and the presence or absence of inverted terminal repeats (ITRs) and PLA2 (1, 2). Iteradensoviruses have a 5-kb genome with ITRs, a monosense genome organization, and PLA2 activity. Four previously known iteradensoviruses are those from *Bombyx mori* (BmDV) (13), *Casphalia extranea* (CeDV) (14), *Dendrolimus punctatus* (DpDV) (15), and *Helicoverpa armigera* (HaDV2) (16). The reported sequence of HaDV2 is incomplete, differs significantly from the sequences of the others, and remains to be confirmed. Additionally, we isolated iteradensoviruses from *Papilio polyxenes* (black swallowtail butterfly) (PpDV) (17), *Sibine fusca* (oil palm pest) (SfDV) (18), and the monarch butterfly (DpIDV) (19) (Fig. 1A). Based on NS1 identities, the taxons of these viruses have been defined as lepidopteran iteradenovirus (LI) 1 with BmDV, the LI 2 species with CeDV, SfDV, and DpIDV, the LI 3 species with DpDV, the LI 4 species with PpDV, and the LI 5 species with HaDV2 (1). NS1 and NS2 proteins of the Chinese isolates (LI 3 and LI 5) had <40% identity scores with other iteradensoviruses. In contrast, except for HaDV2, the VP proteins of iteradensoviruses all have similar identities of 70 to 80%, independent of the species.

Infectious clones were created and sequenced for BmDV (pIN919), CeDV (pSMART-CeDV), PpDV (pCR2.1-PpDV), SfDV (pBlue-SfDV), and DpIDV (pBlue-DpIDV) (14, 17–20). The completely sequenced genomes all have typical J-shaped ITRs with lengths ranging from 230 to 271 nucleotides (nt) and terminal hairpins of 159 to 175 nt (Fig. 1B and C). The 5'- and 3'-terminal hairpins occurred in two orientations, “flip” and its reverse-complement orientation, “flop” (Fig. 1C). These were

identical among the three LI 2 viruses and had few differences in PpDV and BmDV. The positions and sizes of the ORFs were virtually identical in all iteradensoviruses (Fig. 2), except for HaDV2 (which had a truncated NS). The typical parvoviral phospholipase A2 motif (8) was present in all VP1s.

Densovirus transcription has not been studied extensively. So far, VP transcripts of different densoviruses with one VP ORF are not spliced but sets of N-terminally extended proteins are generated by leaky scanning (6, 21–23). Among others, densoviruses from *Blattella germanica* (BgDV) and *Acheta domestica* (AdDV) have two VP ORFs that are joined by splicing and generate proteins with alternate N termini in addition to nested N-terminally extended sets by leaky scanning (4, 24). Leaky scanning is also important for the NS proteins of densoviruses from different genera (4, 6, 21–24). However, splicing determines whether NS3 or NS1/NS2 is expressed for the ambisense densoviruses, whereas the two NS genes of the brevidensoviruses have overlapping promoters so that transcripts start at either side of the NS1 start codon, allowing either NS1 or the downstream NS2 to be expressed (21–24). Transcription of viruses in the *Iteradensovirus* genus has not yet been studied.

Production of iteradensovirus mRNAs and Northern blotting. A major impediment to iteradensovirus studies has been the lack of suitable insect cells that support their replication. Four different insect cell lines were compared for their efficacy to produce iteradensovirus mRNA (Fig. 1D). LD652, Sf, and Bmn cells (in SF-900 II medium) and mosquito C6/36 cells (in RPMI 1640 medium) were propagated with 5% and 10% fetal bovine serum, respectively. Transfections with the infectious clones of CeDV, BmDV, PpDV, and SfDV were performed using DOTAP liposomal transfection reagent (Roche). PpDV was purified as previ-

Received 12 June 2014 Accepted 24 July 2014

Published ahead of print 30 July 2014

Editor: M. J. Imperiale

Address correspondence to Peter Tijssen, peter.tijssen@iaf.inrs.ca.

Copyright © 2014, American Society for Microbiology. All Rights Reserved.

doi:10.1128/JVI.01719-14

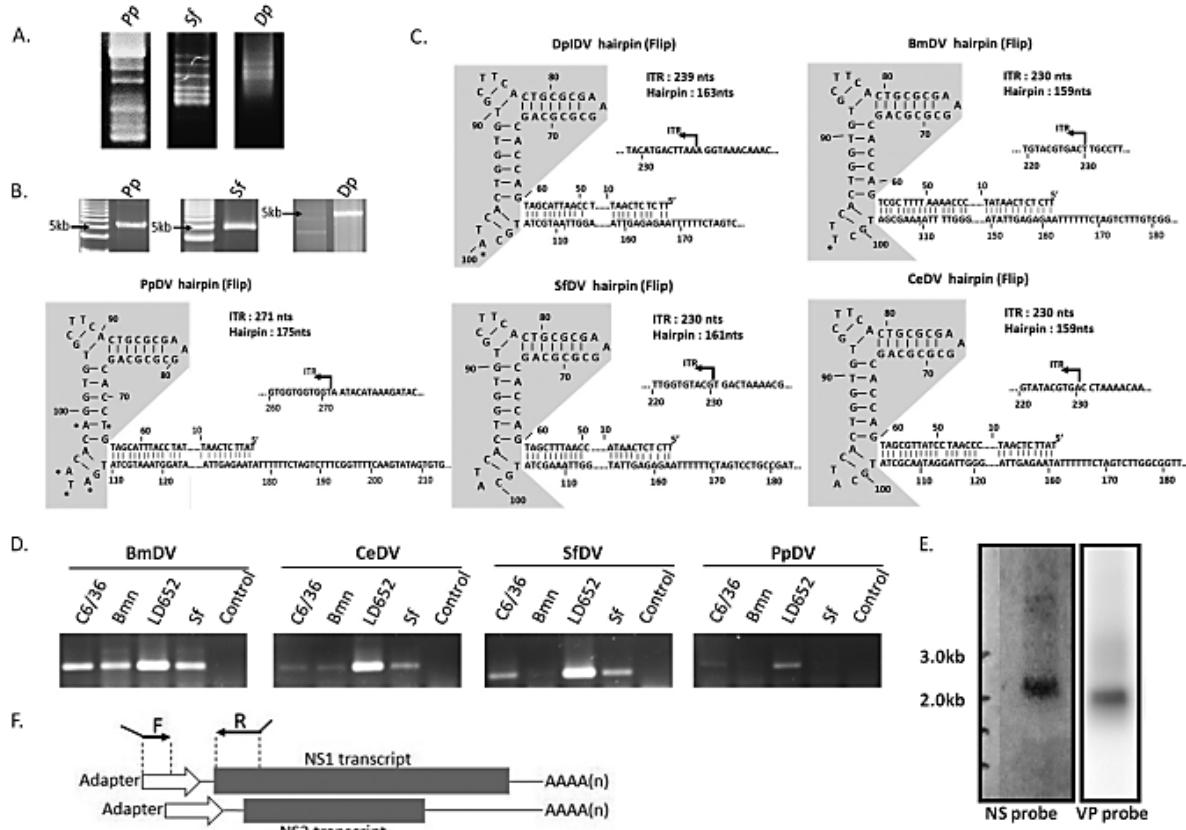


FIG 1 (A) Detection of pathogens by sequence-independent single-primer amplification in larvae from *Papilio polyxenes* (Pp) and *Sibine fusca* (Sf) and in pupae from the monarch butterfly (Dp) on a 2% agarose gel. (B) DNA purification of new iteradenoviruses on 1% electrophoresis agarose gels. The DNA ladder is a 1-kb plus DNA ladder. (C) Inverted terminal repeats (ITRs) of five iteradenoviruses with J-shaped terminal hairpins. The shaded areas in the ITRs occur in two alternative sequences, named "flip" and its reverse complement "flop." Iteradenoviruses' ITR sequences are highly similar to each other. Only a few nucleotides differ and are indicated with an asterisk. (D) Infectious clones of CeDV, BmDV, PpDV, and SfDV were transfected into four different insect cell lines (C6/36, Bmn, Ld652, and Sf), and total RNA was extracted 48 h posttransfection. Gene-specific primers were used for RT-PCR identification. PCR products were checked by electrophoresis in a 1% agarose gel. Total RNAs without reverse transcription were used as a control template for PCR. (E) Northern blotting of PpDV NS and VP transcripts. Both NS and VP yielded single transcript bands. However, further results showed that the 2.3-kb transcripts for NS consisted of two transcripts with almost identical sizes. (F) Scheme of a PCR method for distinguishing NS1 and NS2 transcription starts. A forward primer with ~20 nt of complementary sequence with the RACE adapter at the 3' end and an unspecific extension with restriction sites at the 5' end. The 3' ends of the reverse primers were extended with the sequence complementary to the putative NS1 ATG.

ously described (25). Third-instar *Papilio* larvae, the only host insect available, were fed pesticide-free parsley leaves that had been submerged in a PpDV suspension and served also as a source of PpDV mRNA. Reverse transcription (RT)-PCR results (Fig. 1D) showed that LD652 cells were most suitable for mRNA production of these iteradenoviruses. Total RNA was isolated from *Papilio* larvae 72 h postinfection and 48 h posttransfection from LD652 cells, which had been transfected with SfDV, BmDV, CeDV, or DpIDV infectious clones, by use of the NucleoSpin RNA II kit (Clontech). An extra DNase I treatment was added after RNA extraction with the Turbo DNA-free kit (Ambion). An RT-negative PCR test was included to verify the absence of DNA. Total RNA was subjected to an mRNA purification step using a Poly(A)Purist kit (Ambion). Northern blots were obtained as described previously (21). RNA probes for NS and VP genes were transcribed from NS and VP amplicons obtained by PCR

using primers PpNSRpF700 and PpNSNpR850 for NS and PpVPRpF3200 and PpVPNpR3400 for VP (Table 1). Northern blotting of PpDV mRNA revealed one 2.3-kb band for the NS gene and one 2.0-kb band for the VP gene (Fig. 1E). Since the NS2 ORF overlapped the N-terminal half of NS1, it is expected to share the same poly(A) motif in the 3' ends of their transcripts.

Transcript mapping. Transcript maps for RNA isolated from PpDV-infected *Papilio* larvae or for RNAs isolated from transfected LD652 cells were established by a FirstChoice RLM rapid amplification of cDNA ends (RACE) kit (Ambion) according to the supplier's instructions (see primers in Table 1). All RACE PCR products, cloned into pGEM-T or pGEM-T Easy vectors (Promega), were sequenced by Sanger's method and demonstrated that 5' ends of NS2 transcripts started at the conserved CAGT sites downstream of the expected NS1 ATG (Fig. 2A).

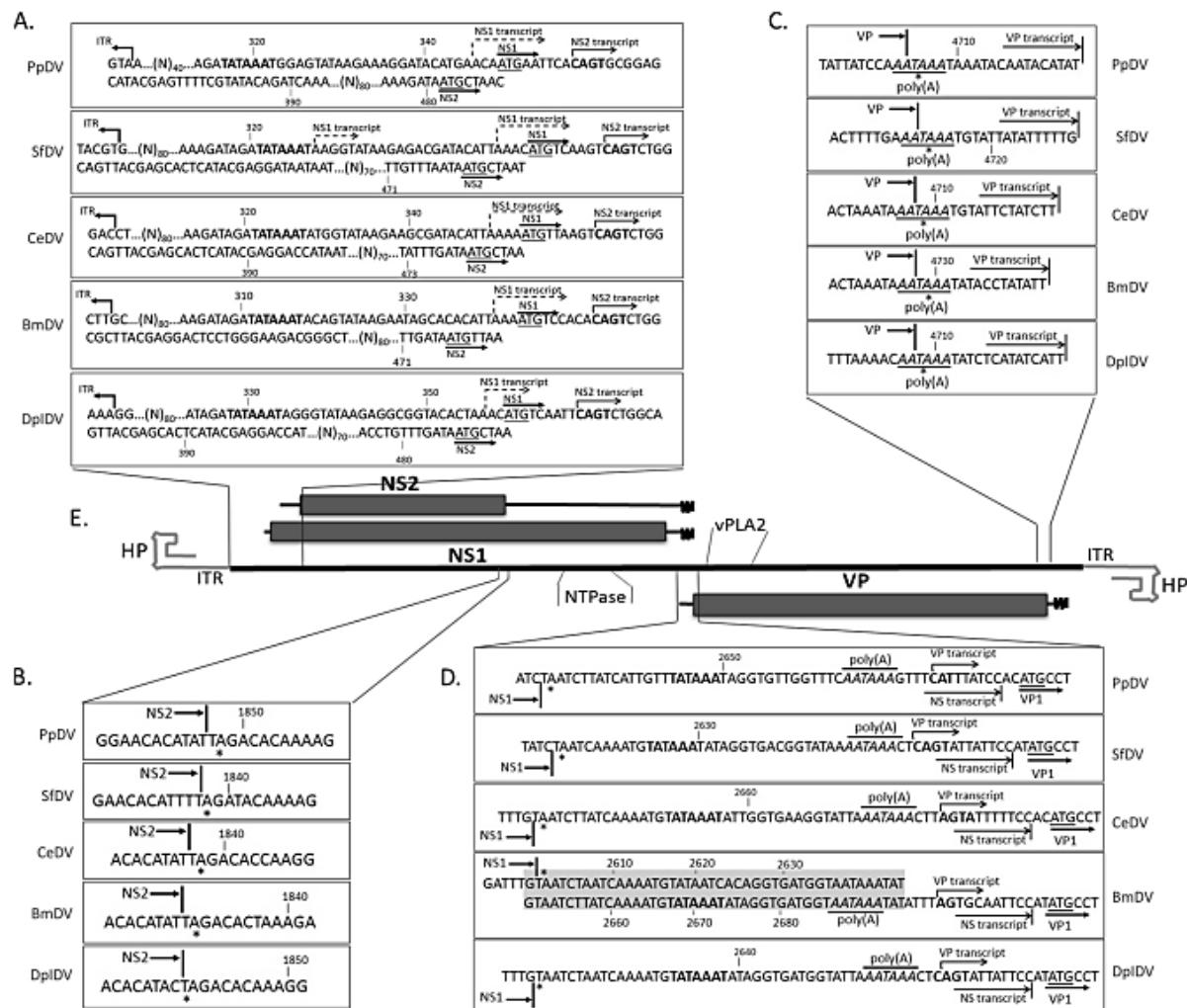


FIG 2 Genome organization and transcription profiles of five iteradenoviruses. (A) The 5' ends of NS transcripts and the ATG sequences of NS1 and NS2 ORFs are underlined. CAGT motifs of NS2 transcription starts are indicated in bold. NS1 transcription starts are marked with dashed arrows. (B) Stop codons of NS2 are indicated with an asterisk. (C) Ends of VP translation and 3' ends of VP transcripts. (D) The 3' ends of NS transcripts and 5'-end starts of VP transcripts. Poly(A) motifs of NS transcripts are in italics. Translation starts of VP are underlined. (E) The genome organization of iteradenoviruses is deduced from the ORFs of all five iteradenoviruses described in this report.

Due to the >40-fold NS2 transcript excess, both in cells and larvae, it was challenging to obtain NS1 transcripts for conventional RACE methods and RNase protection assays. Therefore, a reverse primer (Fig. 1F, labeled R) that extended upstream of the NS2 transcript to the ATG of the putative NS1 start was designed to obtain NS1-specific amplicons. Sequencing of these amplicons revealed the 5'-end start of the NS1 transcripts (Fig. 2A). No differences were observed between mRNA from transfected cells and that from PpDV-infected larvae. The 5' ends of NS1 transcripts usually started 2 to 4 nt upstream of the NS1 ATG (Fig. 2A). Unexpectedly, 5'-untranslated regions of either 4 or 26 nt were obtained by RACE experiments for SfDV NS1 transcripts. NS1 and NS2 3' ends shared the same poly(A) motif downstream of the TATA box of VP transcripts. The BmDV NS transcript poly(A)

was also located downstream of the TATA box of VP transcript, even though this virus contained an intergenic direct repeat of 45 nt between NS1 and VP (Fig. 2B).

All VP transcripts had short 5'-untranslated regions, and the transcripts started only 10 to 15 nt upstream of the first ATG (Fig. 2D) that, moreover, had unfavorable Kozak sequences promoting a leaky scanning translation mechanism for VPs as for ambidenoviruses (6, 22, 23, 26, 27). Previously, we expressed the VP gene of BmDV to generate viruslike particles in order to obtain the three-dimensional structure by X-ray crystallography (28). Poly(A) motifs of the 2.0-kb VP transcripts, determined by 3'-end RACE, overlapped with stop codons of VP proteins for all iteradenoviruses. It would be interesting to obtain transcript maps for LI 3 and 5 viruses as well.

TABLE 1 Primers

Primer ^a	Sequence (5'→3') ^b	Purpose
Nbam24	AGGCAACTGTGCTATCCGAGGGAG	SISPA adapter and primer
NCsp11	TACTCCCTCGG	SISPA adapter and primer
Pa140	GTT GTG GTG GTA TAT ATT GAG AAT AT	Highly conserved primer in iteradensoviruses' ITR
Pp5NS-1	GTCAAGCACTTTCCTGAGGTAA	5'-RACE primer for PpDV NS
Pp3NS2-1	ATAGAACAGAGCCAATACACATGG	3'-RACE primer for PpDV NS2
Pp3NS1-1	ATGCCCATATGGGGTGCTACAGGT	3'-RACE primer for PpDV NS1
Pp5VP-1	CTCCTGTACGTGGATCAACTTGTGC	5'-RACE primer for PpDV VP
Pp3VP-1	GTAATGTCATGACTCCTATACCAGGT	3'-RACE primer for PpDV VP
Sf5NS-1	GTTTCTTCGAAGCTCAATGGAGTC	5'-RACE primer for SfDV NS
Sf3NS2-1	AGCCAGCAATCGAGCCAGTCGGAAG	3'-RACE primer for SfDV NS2
Sf3NS1-1	AACATACAATGCCAATATGGGGTGC	3'-RACE primer for SfDV NS1
Sf5VP-1	TGCTGCTGTATCCATAGGTACTTC	5'-RACE primer for SfDV VP
Sf3VP-1	ATGACTCCAATTCCAGGTTAGAA	3'-RACE primer for SfDV VP
Bm5NS-1	GTTCCATTGTTGATAACTGACAGGG	5'-RACE primer for BmDV NS
Bm3NS1-1	CAGTACTTGCATTTTATTGGAACACTG	3'-RACE primer for BmDV NS2
Bm3NS2-1	GGCCAACATGGCCCTTGGAACAC	3'-RACE primer for BmDV NS1
Bm5VP-1	CGCTTGCCTCTAGCTCTATTAAATC	5'-RACE primer for BmDV VP
Bm3VP-1	GCCTAAGTTATGATGGATTGTAAC	3'-RACE primer for BmDV VP
Ce5NS-1	GCTCAATGGAGTCGTAGAACATTIC	5'-RACE primer for CeDV NS
Ce3NS1-1	CAATGAATATGACCTACAGCGAAC	3'-RACE primer for CeDV NS2
Ce3NS2-1	CCACATACCTCACGCCCTGCCAA	3'-RACE primer for CeDV NS1
Ce5VP-1	CTCTCCAACATCAGTGGCTCC	5'-RACE primer for CeDV VP
Ce3VP-1	GCTTAACTGCAACTAAATAT	3'-RACE primer for CeDV VP
Dpp5NS-1	GCTGAGAAITCCTCGGTTAAGTATTIC	5'-RACE primer for DpIDV NS
Dpp3NS-1	GGAAAGCCAACGTTCAAGAACACCTG	3'-RACE primer for DpIDV NS
Dpp5VP-1	CTCGTCTAGCTGTTCTCGCTCTATTG	5'-RACE primer for DpIDV VP
Dpp3VP-1	CGAAGACAATTCCCTGTTAACTGC	3'-RACE primer for DpIDV VP
PpNSRpF700	GCTTATCAACAAATGAACTATGC	NS probe primer for PpDV
PpNSNpR850 (+T7 sequence)	CAAATTAATACGACTCACTATAGGCTCGGTATTGATGGAGTCG	NS probe primer for PpDV
PpVPRpF3200	GCTCCTCGCTTAATCACATCATC	VP probe primer for PpDV
PpVPNpR3400 (+T7 sequence)	CAAATTAATACGACTCACTATAGGGAGAGCTGTGTAATCTCTG	VP probe primer for PpDV
pIZ5-SfNS1FK	CGGGGTACCGTATAAGAGACGATACTTAAACATG	pIZT/V5-SfNS cloning primer
pIZ5-SfNS2RX	GCTCTAGACCAAATGTTAAAGGGGCCATG	pIZT/V5-SfNS cloning primer
pIZ5-SfNS2FA	CCACCGGTATAACAAAAGATTGGCTAGAAATG	pIZT/V5-SfNS cloning primer
pIZ5-SfNS1RAha (+ HA tag)	CCACCGGT <u>A</u> GCGTAATCTGGAAACATCGTATGGTAGGAGATATC GATAACATACTTATC	pIZT/V5-SfNS cloning primer
SfNS1mATGFn	GAGGATACATTAACACCTCAAGTCAGTCTGGCAG	pIZT/V5-SfNS mutagenesis primer
SfNS1mATGRn	CTGCCAGACTGACTTGGAGGTGTTAATGTATCGTC	pIZT/V5-SfNS mutagenesis primer
SfNS2mATGFn	CTCATTTGTTAAACCTTAATACTACTCTGGTG	pIZT/V5-SfNS mutagenesis primer
SfNS2mATGRn	CACCACTGTTAGGCTTATTAAACAAATGAG	pIZT/V5-SfNS mutagenesis primer
pIZ5-V5muFn	CCCGCGGTTCGAGAGCAAGCCATCCCAAACCCCTCTCTCGG	pIZT/V5-SfNS mutagenesis primer
pIZ5-V5muRn	CCGAGGAGAGGGTTGGGATAGGCTTGCCTCGAACCGGGGG	pIZT/V5-SfNS mutagenesis primer

^a In the primer names, "R" and "F" represent reverse and forward primers, respectively. The numbers in the primer names indicate the 5' end of the primer sequence in virus.

^b The T7 and HA tag sequences are underlined. The first two primers are intended for sequence-independent single-primer amplification (SISPA) (see Fig. 1A).

Expression of nonstructural proteins NS1 and NS2. The pIZT/V5-His insect cell expression vector was used since it contained an OpIE2 promoter for constitutive expression of the NS gene, the Zeocin resistance gene for colony selection, and the cycle 3 green fluorescent protein (GFP) gene for cell line transfection efficiency detection. A V5 epitope was added to the C terminus of NS2, and an HA epitope was added to the C terminus of NS1 (Fig. 3A). An amplicon from NS1 ATG to the NS2 stop codon with CC added to the reverse primer was cloned into the KpnI and XbaI cloning sites of the pIZT/V5-His vector to obtain an in-frame NS2 gene/V5 epitope construct. A second amplicon from the NS2 stop codon to the NS1 stop codon was obtained by PCR with an HA tag added to the reverse primer and a mutation in the NS2 stop codon in the forward primer (Table 1). ATG of the NS1 and NS2 genes

was substituted for ACC (Table 1, primers). To test the expression of SfDV NS1 and NS2 genes by their own promoter, the OpIE2 promoter of the vector was replaced by a fragment between the BspHI and SpeI sites of the viral genome (nt 193 to 750) (Fig. 3A). Two additional mutants were created with the TATA box changed to GAGA and the CAGT motif changed to TTGT. LD652 cells in 24-well plates were transfected with 0.5 μg DNA and 5 μl DOTAP reagent. Cell fixation, indirect immunofluorescence (IF) with 1:200 of V5 antibody (Invitrogen) for staining of NS2 or 1:50 of HA antibody (Santa Cruz) for staining of NS1, and DNA staining with Hoechst 33258 were done as described elsewhere (29).

In comparison to positive controls, NS2 protein expression could be detected only when NS1 ATG was mutated, whereas after NS2 ATG mutant transfection, only NS1 protein expression could

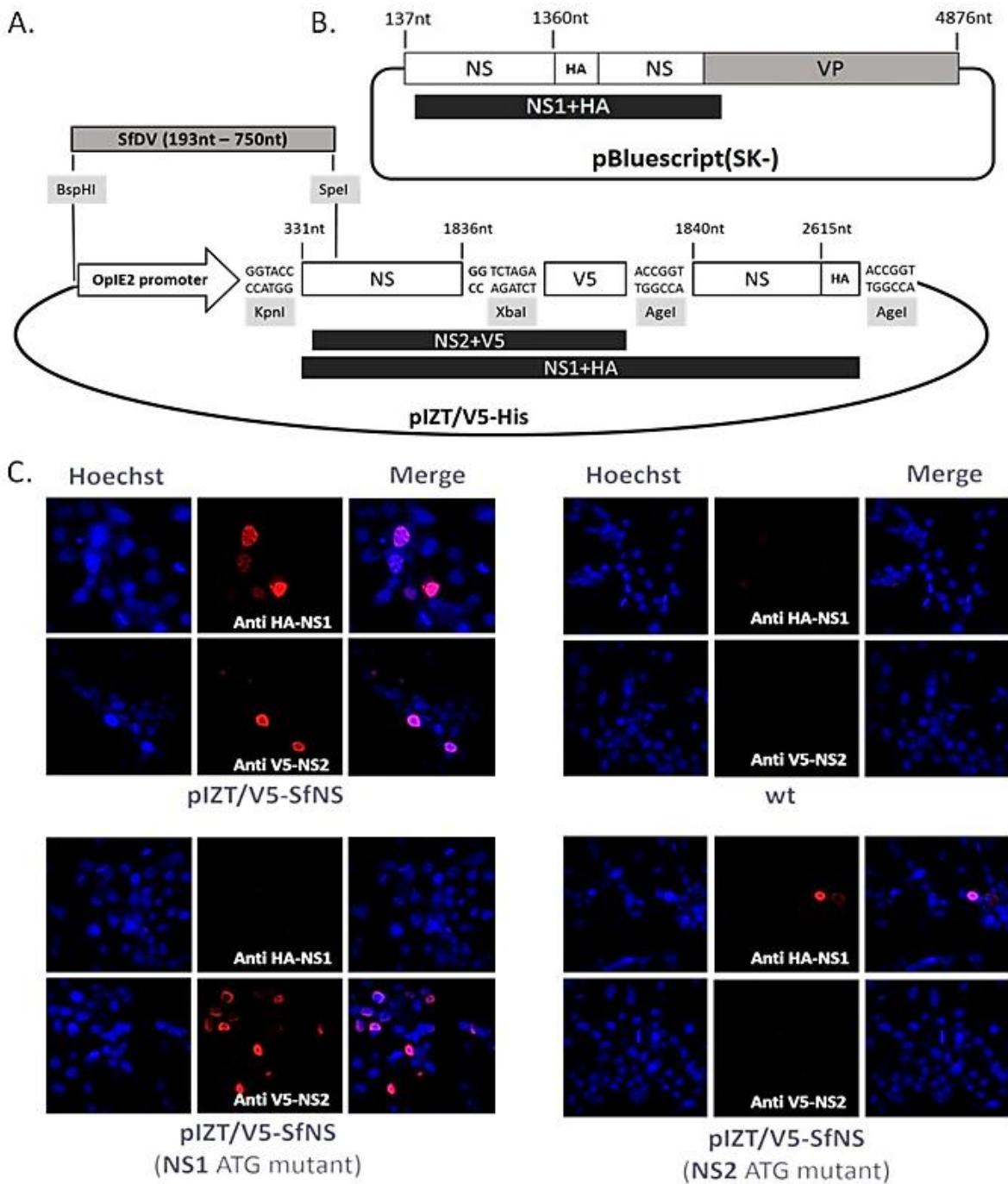


FIG 3 (A) SfDV NS gene expression in the insect expression vector pIZT/V5-His was studied after the addition of HA and V5 tags in frame to NS1 and NS2, respectively. A fragment from nt 331 to 1836, containing the NS2 and N-terminal half of NS1, of the SfDV genome was first cloned between the KpnI and XbaI sites of the pIZT/V5 multiple-cloning site (MCS), CC in bold was added to have the NS2 in frame with V5 epitope. After the V5 sequence, the fragment from nt 1840 to 2615 of the SfDV genome, containing the remainder of NS1, was cloned into the AgeI site and an HA tag sequence was added in frame with NS1 to the 3' end of the insert. To investigate the NS promoter element, the OptIE2 promoter of vector was replaced in alternate constructs by a fragment from nt 193 to 750 of the SfDV genome, between the BspHI and Spel restriction sites. (B) Scheme for introducing an HA tag into the SfDV genome clone to be in frame with the NS1 ORF. SfDV genome sequence from nt 137 to 4876 (lacking the hairpins) was cloned into pBluescript(SK-) vector. An HA tag was added at nt 1360 of the virus genome. Introducing a stop codon in frame immediately downstream of the NS1 start codon blocked NS1 expression, confirming the position of the NS1 translation start. (C) Detection of NS1 and NS2 gene expression in LD652 cells after 48 h posttransfection by immunofluorescence using the HA tag antibody (for NS1) and V5 antibody (for NS2) and an Alexa Fluor 568 goat anti-mouse secondary antibody. pIZT/V5-SfNS constructs allowed detection of both NS1 and NS2 (not possible with the wild type). Mutating the ATG of either the NS1 or NS2 gene blocked their expression.

be detected (Fig. 3C). IF results showed that after transfection of LD652 cells with plasmids containing the original presumed viral promoter element, only NS2 protein expression was detected. After introduction of a stop codon after NS2 ATG or mutation of the CAGT of NS2 transcript, NS2 expression became undetectable. Mutation of the TATA box did not change NS2 expression, which may mean that this TATA box may not be an essential element of the NS2 gene promoter. The viral genome fragment that replaced the OplE2 promoter may not contain the complete NS1 promoter, since NS1 could not be expressed in LD652 cells after transfection. After introduction of an HA tag in frame with NS1 into the SfDV whole-genome clone (except the hairpins) (Fig. 3B), the plasmid expressed NS1 in LD652 cells after transfection. However, NS1 became undetectable after introduction of a stop codon after NS1 ATG, confirming its authenticity as the NS1 start codon and its location upstream of the NS2 transcription start. Combined, these results indicate that iteradenoviruses have, like brevidensoviruses (21), overlapping NS gene promoters that are responsible for different transcript starts and likely dictate the relative amounts of these transcripts.

ACKNOWLEDGMENTS

This work was supported by a Natural Sciences and Engineering Research Council of Canada (NSERC) grant to P.T. Q.Y. acknowledges support from a scholarship from the People's Republic of China and tuition waivers from INRS-IAF.

REFERENCES

- Cotmore SF, Agbandje-McKenna M, Chiorini JA, Mukha DV, Pintel DJ, Qiu J, Soderlund-Venermo M, Tattersall P, Tijssen P, Gatherer D, Davison AJ. 2014. The family Parvoviridae. *Arch. Virol.* 159:1239–1247. <http://dx.doi.org/10.1007/s00705-013-1914-1>.
- Tijssen P, Agbandje-McKenna M, Almendral JM, Bergoin M, Flegel TW, Hedman K, Kleinschmidt JA, Li Y, Pintel DJ, Tattersall P. 2011. Parvoviridae, p 375–395. In King AMQ, Adams MJ, Carstens E, Lefkowitz EJ (ed), *Virus taxonomy: classification and nomenclature of viruses: ninth report of the International Committee on Taxonomy of Viruses*. Elsevier, San Diego, CA.
- Gudenkauf BM, Eaglesham JB, Aragundi WM, Hewson I. 2014. Discovery of urchin-associated densoviruses (family Parvoviridae) in coastal waters of the Big Island, Hawaii. *J. Gen. Virol.* 95:652–658. <http://dx.doi.org/10.1099/vir.0.060780-0>.
- Kapelinskaya TV, Martynova EU, Schal C, Mukha DV. 2011. Expression strategy of densoenceleosis virus from the German cockroach, *Blattella germanica*. *J. Virol.* 85:11855–11870. <http://dx.doi.org/10.1128/JVI.05523-11>.
- Cotmore SF, Tattersall P. 2006. Structure and organization of the viral genome, p 73–94. In Kerr JR, Cotmore SF, Bloom ME, Linden RM, Parrish CR (ed), *Parvoviruses*. Oxford University Press, London, United Kingdom.
- Tijssen P, Bando H, Li Y, Jousset FX, Zadori Z, Fediere G, El-Far M, Szelei J, Bergoin M. 2006. Evolution of densoviruses, p 55–68. In Kerr JR, Cotmore SF, Bloom ME, Linden RM, Parrish CR (ed), *Parvoviruses*. Hodder Arnold, London, England.
- Chapman MS, Agbandje-McKenna M. 2006. Atomic structure of viral particles, p 107–123. In Kerr JR, Cotmore SF, Bloom ME, Linden RM, Parrish CR (ed), *Parvoviruses*. Hodder Arnold, London, England.
- Zadori Z, Szelei J, Lacoste MC, Li Y, Gariepy S, Raymond P, Allaire M, Nabi IR, Tijssen P. 2001. A viral phospholipase A2 is required for parvovirus infectivity. *Dev. Cell* 1:291–302. [http://dx.doi.org/10.1016/S1534-5807\(01\)00031-4](http://dx.doi.org/10.1016/S1534-5807(01)00031-4).
- Canaan S, Zadori Z, Ghomashchi F, Bollinger J, Sadilek M, Moreau ME, Tijssen P, Gelb MH. 2004. Interfacial enzymology of parvovirus phospholipases A2. *J. Biol. Chem.* 279:14502–14508. <http://dx.doi.org/10.1074/jbc.M312630200>.
- Farr GA, Zhang LG, Tattersall P. 2005. Parvoviral virions deploy a capsid-tethered lipolytic enzyme to breach the endosomal membrane during cell entry. *Proc. Natl. Acad. Sci. U. S. A.* 102:17148–17153. <http://dx.doi.org/10.1073/pnas.0508477102>.
- Nüesch JPF. 2006. Regulation of non-structural protein functions by differential synthesis, modification and trafficking, p 275–289. In Kerr JR, Cotmore SF, Bloom ME, Linden RM, Parrish CR (ed), *Parvoviruses*. Hodder Arnold, London, England.
- Cotmore SF, Tattersall P. 2006. A rolling-hairpin strategy: basic mechanisms of DNA replication in the parvoviruses, p 171–188. In Kerr JR, Cotmore SF, Bloom ME, Linden RM, Parrish CR (ed), *Parvoviruses*. Hodder Arnold, London, England.
- Nakagaki M, Kawase S. 1980. Structural proteins of densoenceleosis virus isolated from the silk worm, *Bombyx mori*, infected with flacherie virus. *J. Invertebr. Pathol.* 36:166–171. [http://dx.doi.org/10.1016/0022-2011\(80\)90020-8](http://dx.doi.org/10.1016/0022-2011(80)90020-8).
- Fediere G, Li Y, Zadori Z, Szelei J, Tijssen P. 2002. Genome organization of Casphalia extranea densovirus, a new iteravirus. *Virology* 292:299–308. <http://dx.doi.org/10.1006/viro.2001.1257>.
- Wang J, Zhang J, Jiang H, Liu C, Yi F, Hu Y. 2005. Nucleotide sequence and genomic organization of a newly isolated densovirus infecting *Dendrolimus punctatus*. *J. Gen. Virol.* 86:2169–2173. <http://dx.doi.org/10.1099/vir.0.80898-0>.
- Xu P, Cheng P, Liu Z, Li Y, Murphy RW, Wu K. 2012. Complete genome sequence of a monosense densovirus infecting the cotton bollworm, *Helicoverpa armigera*. *J. Virol.* 86:10909. <http://dx.doi.org/10.1128/JVI.01912-12>.
- Yu Q, Hajek AE, Bergoin M, Tijssen P. 2012. Papilio polyxenes densovirus has an iteravirus-like genome organization. *J. Virol.* 86:9534–9535. <http://dx.doi.org/10.1128/JVI.01368-12>.
- Yu Q, Fediere G, Abd-Alla A, Bergoin M, Tijssen P. 2012. Iteravirus-like genome organization of a densovirus from *Sibine fusca* Stoll. *J. Virol.* 86:8897–8898. <http://dx.doi.org/10.1128/JVI.01267-12>.
- Yu Q, Tijssen P. 2014. Iteradenovirus from the monarch butterfly, *Danaus plexippus plexippus*. *Genome Announc.* 2(2):e00321-14. <http://dx.doi.org/10.1128/genomeA.00321-14>.
- Li Y, Zadori Z, Bando H, Dubuc R, Fediere G, Szelei J, Tijssen P. 2001. Genome organization of the densovirus from *Bombyx mori* (BmDNV-1) and enzyme activity of its capsid. *J. Gen. Virol.* 82:2821–2825.
- Pham HT, Jousset FX, Perreault J, Shike H, Szelei J, Bergoin M, Tijssen P. 2013. Expression strategy of *Aedes albopictus* densovirus. *J. Virol.* 87: 9928–9932. <http://dx.doi.org/10.1128/JVI.01259-13>.
- Fediere G, El-Far M, Li Y, Bergoin M, Tijssen P. 2004. Expression strategy of densoenceleosis virus from *Mythimna loreyi*. *Virology* 320:181–189. <http://dx.doi.org/10.1016/j.virol.2003.11.033>.
- Tijssen P, Li Y, El-Far M, Szelei J, Letarte M, Zadori Z. 2003. Organization and expression strategy of the ambisense genome of densoenceleosis virus of *Galleria mellonella*. *J. Virol.* 77:10357–10365. <http://dx.doi.org/10.1128/JVI.77.19.10357-10365.2003>.
- Liu K, Li Y, Jousset FX, Zadori Z, Szelei J, Yu Q, Pham HT, Lepine F, Bergoin M, Tijssen P. 2011. The *Acheta domesticus* densovirus, isolated from the European house cricket, has evolved an expression strategy unique among parvoviruses. *J. Virol.* 85:10069–10078. <http://dx.doi.org/10.1128/JVI.00625-11>.
- Tijssen P, Tijssen-van der Slikke T, Kurstak E. 1977. Biochemical, biophysical, and biological properties of densoenceleosis virus (paravovirus). II. Two types of infectious virions. *J. Virol.* 21:225–231.
- Huynh OT, Pham HT, Yu Q, Tijssen P. 2012. *Pseudoplusia includens* densovirus genome organization and expression strategy. *J. Virol.* 86: 13127–13128. <http://dx.doi.org/10.1128/JVI.02462-12>.
- El-Far M, Szelei J, Yu Q, Fediere G, Bergoin M, Tijssen P. 2012. Organization of the ambisense genome of the *Helicoverpa armigera* densovirus. *J. Virol.* 86:7024. <http://dx.doi.org/10.1128/JVI.00865-12>.
- Kaufmann B, El-Far M, Plevka P, Bowman VD, Li Y, Tijssen P, Rossman MG. 2011. Structure of *Bombyx mori* densovirus 1, a silk-worm pathogen. *J. Virol.* 85:4691–4697. <http://dx.doi.org/10.1128/JVI.02688-10>.
- Boisvert M, Fernandes S, Tijssen P. 2010. Multiple pathways involved in porcine parvovirus cellular entry and trafficking toward the nucleus. *J. Virol.* 84:7782–7792. <http://dx.doi.org/10.1128/JVI.00479-10>.

The *Acheta domesticus* Densovirus, Isolated from the European House Cricket, Has Evolved an Expression Strategy Unique among Parvoviruses[†]

Kaiyu Liu,^{1,2}§ Yi Li,^{1,2}§ Françoise-Xavière Jousset,³§ Zoltan Zaduri,¹‡ Jozsef Szelei,¹ Qian Yu,¹ Hanh Thi Pham,¹ François Lépine,¹ Max Bergoin,^{1,3} and Peter Tijssen^{1*}

*INRS-Institut Armand Frappier, Université du Québec, Laval, Quebec, Canada*¹; *Central China Normal University, Wuhan 430079, People's Republic of China*²; and *Laboratoire de Pathologie Comparée, Université Montpellier 2, Montpellier 34095, France*³

Received 29 March 2011/Accepted 12 July 2011

The *Acheta domesticus* densovirus (AdDNV), isolated from crickets, has been endemic in Europe for at least 35 years. Severe epizootics have also been observed in American commercial rearings since 2009 and 2010. The AdDNV genome was cloned and sequenced for this study. The transcription map showed that splicing occurred in both the nonstructural (NS) and capsid protein (VP) multicistronic RNAs. The splicing pattern of NS mRNA predicted 3 nonstructural proteins (NS1 [576 codons], NS2 [286 codons], and NS3 [213 codons]). The VP gene cassette contained two VP open reading frames (ORFs), of 597 (ORF-A) and 268 (ORF-B) codons. The VP2 sequence was shown by N-terminal Edman degradation and mass spectrometry to correspond with ORF-A. Mass spectrometry, sequencing, and Western blotting of baculovirus-expressed VPs versus native structural proteins demonstrated that the VP1 structural protein was generated by joining ORF-A and -B via splicing (splice II), eliminating the N terminus of VP2. This splice resulted in a nested set of VP1 (816 codons), VP3 (467 codons), and VP4 (429 codons) structural proteins. In contrast, the two splices within ORF-B (Ia and Ib) removed the donor site of intron II and resulted in VP2, VP3, and VP4 expression. ORF-B may also code for several nonstructural proteins, of 268, 233, and 158 codons. The small ORF-B contains the coding sequence for a phospholipase A2 motif found in VP1, which was shown previously to be critical for cellular uptake of the virus. These splicing features are unique among parvoviruses and define a new genus of ambisense densoviruses.

Insect parvoviruses (densoviruses) belong to the *Densovirinae* subfamily of the *Parvoviridae* and are small, isometric, nonenveloped viruses (diameter, ~25 nm) that contain a linear single-stranded DNA of 4 to 6 kb (2, 3, 27). These viruses can be subdivided into two large groups, those with ambisense genomes and those with monosense genomes. Like vertebrate parvoviruses, all densoviruses have a genomic DNA with hairpins at both ends, often (but not necessarily for all genera) as inverted terminal repeats (ITRs). All densoviruses with ambisense genomes package both complementary strands in equimolecular ratios as single strands in separate capsids (27). The nonstructural (NS) gene cassette is found in the 5' half of one genome strand, and the structural protein (VP) gene cassette is found in the 5' half of the complementary strand. By convention, the genome is oriented so that the NS cassette is found in the left half. Expression strategies of densoviruses often involve (alternative) splicing and leaky scanning translation mechanisms (28). So far, the near-atomic structures of three densoviruses, *Penaeus stylirostris* densovirus (PstDNV),

Bombyx mori densovirus 1 (BmDNV-1), and *Galleria mellonella* densovirus (GmDNV), have been solved (10, 11, 21). The capsid of densoviruses consists of 60 subunits (T=1) of identical proteins that may contain N-terminal extensions not involved in capsid formation but that confer additional functions to the capsid. One of these functions is a phospholipase A2 (PLA2) activity that is required for genome delivery during infection (34). Densoviruses are usually highly pathogenic for their natural hosts (5).

The monosense densoviruses have been classified into 3 uniform genera, i.e., *Iteravirus*, with a 5.0-kb genome, 0.25-kb ITRs, and a PLA2 motif in VP; *Brevidensovirus*, with a 4.0-kb genome, no ITRs but terminal hairpins, and no PLA2 motif; and *Hepanvirus*, with a single member, hepatopancreatic parvovirus, with a 6.3-kb genome also lacking a PLA2 motif and ITRs but with 0.2-kb terminal hairpins (23, 27). In contrast, the ambisense densoviruses have been classified into one uniform genus, *Densovirus*, with a 6-kb genome and 0.55-kb ITRs, and a second genus, *Pefudenvirus*, with only *Periplaneta fuliginosa* densovirus (PfdDNV) as a member, with a 5.5-kb genome, 0.2-kb ITRs, and a split VP gene cassette (2, 26). Ribosome frameshifts have been proposed to connect its VP open reading frames (ORFs) (33). So far, all ambisense densoviruses have an N-terminal PLA2 motif in their largest VP. Some sequenced ambisense densoviruses, e.g., *Myzus persicae* densovirus (MpDNV) (32), *Blattella germanica* densovirus (BgDNV) (18), and *Planococcus citri* densovirus (PcDNV) (25), are as yet unclassified. The ambisense virus *Culex pipiens* densovirus (CpDNV) has a different genome organization for both the NS

* Corresponding author. Mailing address: INRS-Institut Armand Frappier, Université du Québec, 531 Boul. des Prairies, Laval, Quebec, Canada H7V 1B7. Phone: (450) 687-5010, ext. 4425. Fax: (450) 686-5501. E-mail: peter.tijssen@iaf.inrs.ca.

† Supplemental material for this article may be found at <http://jvi.asm.org/>.

‡ Present address: Veterinary Medical Research Institute, H-1143 Budapest, Hungary.

§ K.L., Y.L., and F.-X. J. contributed equally to this work.

† Published ahead of print on 20 July 2011.

and VP proteins and will have to be classified in a different genus (1).

Acheta domesticus densovirus (AdDNV) was isolated from diseased *Acheta domesticus* L. house crickets from a Swiss commercial mass rearing facility (16). The virus spread rapidly and was responsible for high mortality rates, such that the rearing could not be saved. This was the first observation of a densovirus in an orthopteran species. Infected tissues included adipose tissue, the midgut, the hypodermis, and particularly the Malpighi tubules, but the most obvious pathological change was the completely empty digestive cecae (24). The cecae, which flank the proventriculus, are the sites where most enzymes are released and most absorption of nutrients occurs. Feulgen-positive masses were observed in the nuclei of infected cells (16). Commercial production facilities for the pet industry or for research mass rearings of house crickets in Europe are frequently affected by this pathogen. This virus was previously not known to circulate in North America, except for a small epidemic in the Southern United States in the 1980s (22). Beginning in 2009, sudden, severe outbreaks were observed in commercial facilities in Canada and the United States, leading to losses of hundreds of millions of dollars and to an acute crisis in the pet food industry (24). In this study on AdDNV, we observed that over the last 34 years the annual replacement rate was about 2.45×10^{-4} substitution/nucleotide (nt) and that the VP gene cassette consists of two ORFs, a characteristic of the *Pefudenvirus* genus (24).

In the present study, the complete genome and the expression strategy of AdDNV were investigated and showed features not yet described for other densoviruses or vertebrate parvoviruses. The most striking observation was the intricate splicing pattern of its VP ORFs, resulting, in contrast to the case for all parvoviruses studied so far, in unrelated N-terminal extensions of its two largest structural proteins and in the probable production of several supplementary NS proteins from the VP cassette.

MATERIALS AND METHODS

Rearing of crickets. *A. domesticus* L. house crickets were obtained from a commercial supplier and were reared under conditions of about 60% relative humidity, 28°C, and a 16-h-8-h light-dark cycle. Diet conditions and drinking water supply, as well as conditions for perching, hiding, and oviposition, were as described previously (24).

Infection techniques. The visceral cavity of nymphs of about 1.5 to 2 cm was injected with an inoculum consisting of a viral suspension obtained by grinding an infected cricket in 1× phosphate-buffered saline (PBS) plus 2% ascorbic acid, clarifying the mixture by centrifugation for 10 min at 8,000 × g, and filtering it through 450-nm membranes. Mortality was usually 100% within 2 weeks. Alternatively, infection was achieved by feeding with a virus-contaminated diet as previously described (24).

Virus and DNA preparation. Virus was purified as previously described (29). Lysis buffer [300 µl of 6 M guanidine-HCl, 10 mM urea, 10 mM Tris-HCl, 20% Triton X-100, pH 4.4, containing 80 µg/ml poly(A) carrier RNA] and 200 µl sample were mixed and incubated for 10 min at 30°C. The sample was vortexed after adding 125 µl isopropanol, and the DNA was then purified on High-Pure plasmid spin columns (Rochi Molecular Biochemicals) according to the supplier's instructions.

Cloning, mutation analysis, and sequencing of viral DNA. The 1977 isolate of AdDNV was cloned into the pCR-XL-TOPO vector (Invitrogen Life Sciences), using supercompetent Sure 2 *Escherichia coli* cells (Stratagene) at 30°C. Point mutations in the AdDNV genome were generated with a QuikChange site-directed mutagenesis kit (Stratagene), whereas deletion mutants were obtained via the type IIb restriction endonuclease strategy (7). Independent clones were sequenced in both directions by primer walking. The terminal hairpins yielded

compressions that were difficult to sequence; however, inclusion of 1 M betaine (Sigma) and 3% dimethyl sulfoxide (DMSO) or restriction in the hairpin by DraI yielded clean sequence reads. DNAs from subsequent isolates were amplified by PCR and sequenced between the ITRs.

Isolation and characterization of viral RNA. Total RNAs were isolated from 30 mg adipose tissue from infected cricket larvae (2 to 5 days postinfection [p.i.]) and from recombinant baculovirus-infected cells at 48 h p.i. by use of an RNeasy minikit from Qiagen. The DNase I treatment was extended from 15 to 30 min or repeated twice. A PCR test was included to verify the absence of DNA. Total extracted RNA was subjected to mRNA purification using an mRNA isolation kit (Roche).

Northern blots. About 20 to 30 µg total RNA in a 6-µl volume was added to 18 µl buffer (1× MOPS [20 mM morpholinepropanesulfonic acid, 5 mM sodium acetate, 0.5 mM EDTA adjusted to pH 8 with NaOH], 18.5% formaldehyde, 50% formamide), 5 µl loading buffer was added, and the mixture was incubated for 5 to 10 min at 65 to 70°C and separated by electrophoresis on a 1% formaldehyde-agarose gel. Parallel lanes contained RNA size markers (Promega). After migration and washing, RNAs were transferred to positively charged nylon membranes (Roche) by capillary blotting overnight. The blotted membranes were prehybridized with 10 mg/ml herring sperm DNA in 50% formamide before hybridization with ³²P-labeled probes. The probes corresponded to a 1.5-kb BglII-Sall restriction fragment specific for the VP coding sequence and a 0.87-kb Eco47III-DraI restriction fragment specific for NS. Hybridized probes were visualized with a Storm 840 phosphorimager.

Mapping of 5' ends, 3' ends, and introns of viral transcripts. The most probable locations of the transcripts were predicted from the ORFs obtained by sequence analysis. A 3' rapid amplification of cDNA ends (3'-RACE) system was used to characterize the 3' ends of the polyadenylated transcripts, using the RNATag and ADAP primers (Table 1) and PCR (28), whereas the 5' ends were determined with a FirstChoice RLM RACE kit (Ambion) according to the instructions of the supplier. The locations of introns were determined after reverse transcription of the transcripts by use of avian myeloblastosis virus (AMV) reverse transcriptase (Promega) in a final volume of 20 µl for 1 h at 42°C, PCR using internal DNV-specific primers for overlapping regions, and dideoxy sequencing of the amplicons according to standard methods (28).

Promoter activity in AdDNV genome. Promoter regions were amplified by PCR and cloned upstream of the luciferase gene in the pGL3-basic system. The ProNSF and ProNSR primers were used for the NS promoter, the ProVPIF and ProVP1R primers were used for the VP1 promoter, and the PrNSMf and PrNSMr primers were used for the *Mythimna loryi* densovirus (MIDNV) NS promoter (control) (Table 1). Sequencing was performed to confirm the promoter direction. For the assay, Ld652 cells were seeded into wells of 24-well cell culture plates. Each well contained about 0.5 ml of cells at 5×10^5 cells/ml. The cells were cultured overnight. Transfection was performed with 2.5 µl DOTAP reagent and 0.6 µg DNA in 15 µl HEPES, and the mixture was added to 245 µl medium (without antibiotic or fetal bovine serum [FBS]) per well. Cells were harvested at 48 or 60 h posttransfection, washed twice with PBS, and resuspended in 100 µl of Bright-Glo lysis buffer (Promega). Cell lysates were quickly centrifuged to remove cell debris, and 25-µl aliquots of the cell extract were used to determine luciferase activity according to the instructions for a luciferase assay system (Promega).

Expression of structural proteins and analysis of VP ORFs by use of a baculovirus system. The potential VP coding sequences (see below) were cloned into the *Autographica californica* nuclear polyhedrosis virus (AcNPV) downstream of the polyhedrin promoter by use of the Bac-To-Bac baculovirus expression system (14) (Invitrogen) via the pFastBac1 and pFastBacHT vectors according to the supplier's instructions. In constructs involving expression of VP1, the initiation codon had to be moved closer to the start of the transcript. For this purpose, an EcoRI site was introduced 100 bp upstream of the multiple cloning site (MCS), using the pFECRF and pFECRIR mutation primers (Table 1), followed by removal of the small EcoRI fragment between the new and MCS EcoRI sites. Inserts were generated by PCR (28) with the primers given in Table 1, using the wild-type (wt) template or a template in which intron II (see below) splicing sites had been mutated. The forward primer with an EcoRI site was either AdATG1B, which coincided with the initiation codon of VP1, or an equivalent in which the initiation codon ATG was mutated to ACC (AdmATG1B), and the reverse primer AdIHAR, containing an XbaI site, was used (Table 1). All pFastBac recombinant constructs were verified by sequencing.

Protein analysis by SDS-PAGE, Western blotting, and N-terminal amino acid sequencing. Capsid proteins were analyzed by SDS-PAGE (13), using the structural proteins of *Junonia coenia* densovirus (JcDNV) or broad-range standards (Bio-Rad) as size markers. Expressed proteins were analyzed by Western blotting (28, 30), using polyvinylidene difluoride (PVDF) membranes and Roche blocking reagent. For amino acid sequencing, structural proteins from AdDNV

TABLE 1. PCR primers used in this study

Primer	Sequence ^a	Position (nt) in AdDNV	Target or use
AdVPR	TTTGTGCAATCCATAATAGTAC	2610–2633	Near 3' end of VP mRNA
NAdR	gettagatCATCTGAACGTTTACCAACT	3892–3915	Just upstream of VP4
Adsp	tggaaatcCAGTTCTGGATGAGG	4362–4380	19–37 nt into VP2
Adsvp	gcTACCAAGAAATCCGTAAATGACA	4546–4534/4403–4393	Small intron splice
Ad3s	CGTAGTACTGATACTTTTATT	4435–4412	End of ORF-B (TGA)
Ads	gcCCTCAACACCTAAAAACGTGAGTACTGA	4453–4424	End of ORF-B (TGA)
Ad6s	CCTAAAACAGTGA <u>TGA</u>	4444–4424	End of ORF-B (TGA)
Adl	gccGACGTATTGGGACCTGTATATCCT	4477–4451	End of ORF-B (–4 aa)
Adm	gcCCTGTATATCCTCAACACCTAA	4462–4439	End of ORF-B (–4 aa)
AdgF	tcggaaatcATGTC <u>TGGCGT</u> CTTACA	5230–5213	Start of VP1 (ORF-B)
RNAtag	gggtctagactcgagT ₁₇	Poly(A) sequence	3'/5'-RACE (first round)
ADAP	gggtctagactcgagT		Subsequent rounds of RACE
ProNSF	acggatt <u>GATAAAGAGCAAGCACCC</u>	109–128	NS1 promoter (KpnI)
ProNSR	gaagat <u>CTGTGCTGGAGGCCTACTGCAGCGAACAC</u>	240–207	NS1 promoter (reverse)
ProVP1R	gaagat <u>AAGACGCCAGAGATTAA</u> ACT	5217–5238	VP1 (BglII site; reverse)
ProVP1F	gtaggta <u>cGATATAAAGAGCAAGCACCA</u>	5323–5300	VP1(KpnI site; forward)
PrNSMf	ACGGTACCGACTATAAATAGAGCTGAGC		MIDNV (forward)
PrNSMr	GAAGATCTATCTGCAATAGATATACTACCA		MIDNV (reverse)
pFECRIF	CGCAAATAATAAGAATTCTACTGTTTCGTAAC		Mutation in pFastBac1
pFECRIR	GTTACGAAAACAGTAGAAATTCTATTATTGCG		Mutation in pFastBac1
Ad1HAR	GCTCTAGATCAAGCGTAATCTGGAACATCGTATGGGTA		VP products (reverse primer [HA])
	TTTTGTTGCAATCCATAATA		
AdATG1B	ggaattcATGTC <u>CGT</u> TTACAGATCTCAC		
AdmATG1B	ggaat <u>cACCTCCGT</u> TTACAGATCTCAC		

^a The AdDNV sequence is shown in capital letters. Underlined nucleotides indicate stop codons (e.g., TGA) or restriction sites (e.g., ggtaac [KpnI]).

were separated by SDS-PAGE on 10% polyacrylamide gels and were electroblotted onto nitrocellulose membranes (Westran, Schleicher & Schuell, Keene, NH) and sequenced according to the method of Matsudaira (15).

MS. Expressed proteins from baculovirus constructs and native proteins from the virus were analyzed by mass spectrometry (MS) after separation by SDS-PAGE. The proteins, dissociated with 2% SDS at 95°C for 5 min, were run in a 10% acrylamide gel (13). The protein bands were cut from the gel and destained with water-sodium bicarbonate buffer and acetonitrile. Each protein was reduced with dithiothreitol (DTT) and alkylated with iodoacetamide prior to in-gel digestion with trypsin (8). The tryptic peptides were eluted from the gel with acetonitrile containing 0.1% trifluoroacetic acid. The tryptic peptides were then separated on an Agilent Nanopump instrument using a C₁₈ Zorbax trap and an SB-C₁₈ Zorbax 300 reversed-phase column (150 mm × 75 μm; 3.5-μm particle size) (Agilent Technologies, Inc.). All mass spectra were recorded on a hybrid linear ion trap-triple-quadrupole mass spectrometer (Q-Trap; Applied Biosystems/MDS Sciex Instruments, CA) equipped with a nano-electrospray ionization source. The analysis of MS-MS data was performed with Analyst software, version 1.4 (Applied Biosystems/MDS Sciex Instruments, CA). MASCOT (Matrix Science, London, United Kingdom) was used to create peak lists from MS and MS/MS raw data.

Nucleotide sequence accession number. The AdDNV sequence is available in the GenBank database under accession number HQ827781.

RESULTS

AdDNV infection of *A. domesticus*. AdDNV is a frequent cause of epizootics in commercial or research mass rearing facilities for house crickets in Europe. The highest mortality is observed in the last larval stage and in young adults. These crickets die slowly over a period of several days; although they appear healthy, they lie on their backs and do not move. The guts of infected *A. domesticus* crickets that are still alive and no longer move are almost always completely empty. Beginning in September 2009, mass epizootics have also occurred in rearing facilities throughout North America.

DNA sequence and organization of AdDNV isolates. Three full-length genomic clones in the pCR-XL-TOPO vector,

namely, pAd22, pAd25, and pAd35, were obtained from the 1977 AdDNV isolate. Both strands of the viral genomes were sequenced (for the full annotated sequence, see Fig. S1 in the supplemental material). Nucleotide substitutions in more recent isolates have been reported elsewhere (24). The total length of the genome was 5,425 nt and contained ITRs of 144 nt, of which the distal 114 nt could fold into a perfect I-type palindromic hairpin (Fig. 1A, B, and D). The side arms in the typical Y-shaped terminal palindromes of many parvoviruses were missing in the case of AdDNV.

Both complementary strands contained large ORFs in their 5' halves; one strand had 3 large ORFs (ORFs 1 to 3), 2 of which were overlapping, and its complementary strand had 2 large ORFs (ORF-A and -B) (Fig. 1A). ORFs 1 to 3 potentially code for proteins consisting of 576, 286, and 213 amino acids (aa), respectively, whereas ORFs A and B potentially code for proteins of 597 and 268 aa, respectively. nBLAST analysis (<http://www.ncbi.nih.gov>) of the 5 ORFs revealed that the ORF-1 product is a homologue of densovirus NS1 proteins, the ORF-2 product is a homologue of densovirus NS2 proteins, the ORF-A product is a homologue of densovirus VP proteins, the ORF-B product is a homologue of parvoviral phospholipase A2 proteins (N-terminal sequence of VP1), and the ORF-3 product does not have homologous proteins. NS1 ORF-1 can also be recognized by the presence of rolling circle replication and Walker A and B motifs, and the VP ORF can be recognized by the presence of a PLA2 motif (Fig. 1A; see Fig. S1 in the supplemental material). Since the convention for all parvoviruses is to have the genes coding for the nonstructural proteins in the left half of the genome, it was decided to define the strand of the ambisense AdDNV genome con-

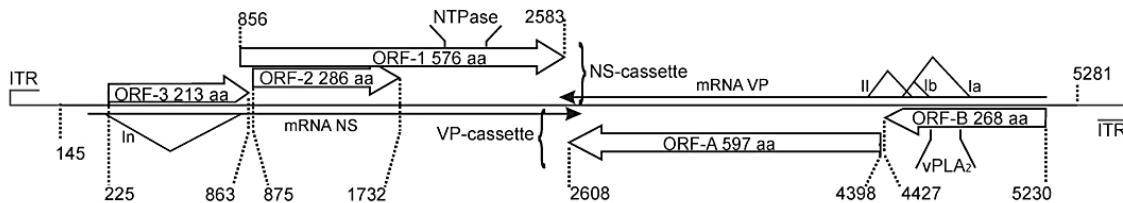
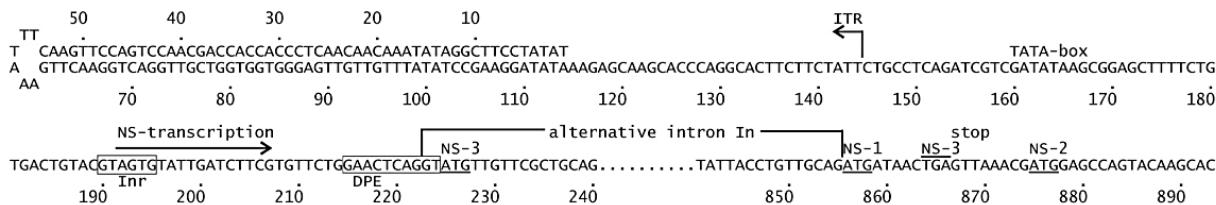
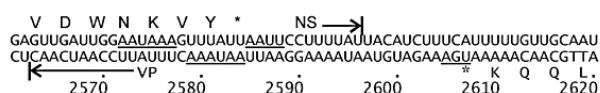
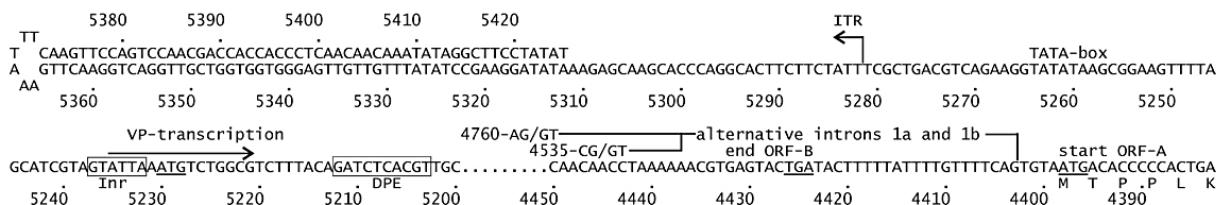
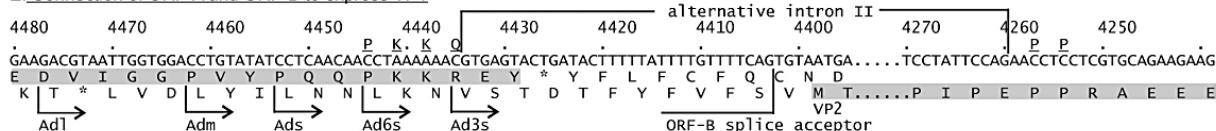
A. Genome organization: ITRs, ORFs and introns**B. Left ITR, NS promoter elements and NS transcription****C. Transcription/translation ends and overlap of mRNAs****D. Right ITR, VP promoter elements, VP transcription and ORFs-A and B****E. Connection of ORF-A and ORF-B to express VP1**

FIG. 1. Genome organization of AdDNV. (A) Overview of genome organization, positions, and sizes of ITRs, ORFs, and introns. The Intron in the NS mRNA, between nt 223 and nt 855, occurs in about half of the NS transcripts. Ia (nt 4403 to 4758), Ib (nt 4403 to 4533), and II (nt 4260 to 4434) are introns that occur in alternative VP transcripts. vPLA2 indicates the position of the viral phospholipase A2 motif. (B) Left ITR and regulation of production of NS transcripts. NS transcripts start at 192-A and yield NS3 from 225-A. However, a fraction of these transcripts are spliced just upstream of this start codon (intron In), leading to translation of NS1 from 856-A (AUG with a poor initiation environment) and, through leaky scanning, of NS2 from 875-A. Inr and DPE are promoter elements. (C) Like the case for all members of the *Denovirus* genus, the 3' ends of AdDNV NS and VP transcripts overlap in the middle of the genome. The stop codons and AATAAA motifs are underlined. (D) Right ITR, VP transcription sites, and splicing in ORF-B on the complementary strand. Transcription starts at nt 5235, and VP1 initiation is at nt 5230. The short 5'-UTR predicts an inefficient initiation (leaky scanning) and could be responsible for the production of a nested set of N-terminally extended viral proteins. However, removal of either of the two alternative introns in ORF-B (Ia or Ib) did not connect the exons in ORF-B and ORF-A in frame, so only nonstructural proteins could be produced from nt 5230 and VP2 could be produced directly from the first AUG in ORF-A when this splicing occurred. (E) An alternative intron II, which is mutually exclusive with introns Ia and Ib because the ORF-B splice acceptor is removed, connects ORF-B and ORF-A (both shaded) in frame so that VP1 can be produced from nt 5230. The VP1 sequence around the splicing site is underlined and shown above the nt sequence.

taining the ORFs for the NS genes as the plus strand so the genes would be located similarly.

SDS-PAGE revealed that the capsid is composed of 4 structural proteins with estimated molecular masses ranging from 43 to 110 kDa (Fig. 2A), although a fifth protein may arise during purification, probably due to proteolysis (not shown). Attempts to obtain N-terminal sequences failed for VP1, VP3,

and VP4, but the sequence TPPLKPHP(I)(E) was obtained for VP2, which indicated that its translation started at the AUG start codon of ORF-A, at nt 4398 to 4396 (Fig. 1D; see Fig. S1 in the supplemental material), and predicted a molecular mass of 65.3 kDa for VP2. ORF-B encoded the PLA2 motif recently identified in the structural proteins of most parvoviruses (4, 6, 28, 34) but was too small to code for a VP1 of 110 kDa as

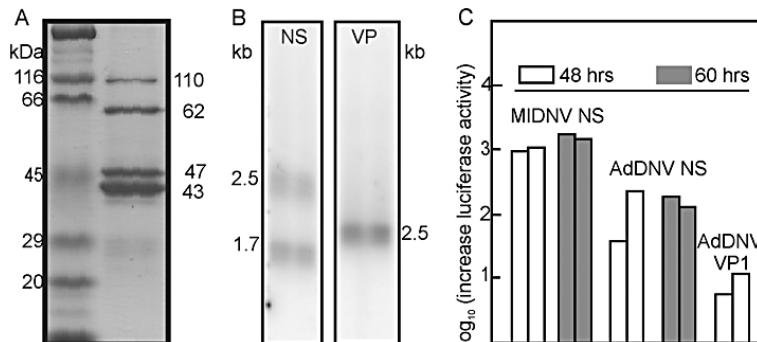


FIG. 2. (A) SDS-PAGE analysis of structural proteins of AdDNV. Lane 1, Bio-Rad broad-range standard proteins as markers; lane 2, four proteins of AdDNV (VP1 to -4, in decreasing size). The estimated masses corresponded reasonably with the sequence-predicted masses (88, 65, 51, and 47 kDa, respectively). (B) Northern blotting of NS and VP transcripts. The two NS transcripts corresponded to a spliced and an unspliced form (see the text for further details). The single VP band consisted of at least three forms of almost identical size. (C) Promoter activities of the predicted NS and VP promoters in two independent experiments using the pGL3 vector, lepidopteran LD cells, and the Promega luciferase assay. The NS promoter was assayed at both 48 and 60 h. The NS promoter of MIDNV, a lepidopteran virus, was used as a control in the lepidopteran cells.

estimated by SDS-PAGE (Fig. 2A). Therefore, splicing of ORF-A and ORF-B seemed necessary to code for the largest AdDNV structural protein. Since the N-terminal coding region of ORF-A before its first ATG overlapped with the C-terminal coding region of ORF-B (Fig. 1E; see Fig. S1 in the supplemental material), an unspliced transcript could also code for VP1 by a ribosome frameshift in the ORF-A-ORF-B overlapping region, as suggested for PfDNV (33). These hypotheses were investigated further by transcript mapping and expression studies.

Northern blotting and mapping of viral transcripts. Northern blotting of RNAs obtained from infected *Acheta* larvae revealed two bands of NS transcripts (about 2.5 and 1.8 kb) and one band of VP transcripts (about 2.5 kb) (Fig. 2B). The transcript maps for RNAs isolated from both diseased crickets and recombinant baculovirus-infected cells were established by 5'- and 3'-RACE and are shown in Fig. 1B to D. The 3' termini of the NS and VP transcripts had a 34-nt overlap (Fig. 1C), similar to the situation observed with members of the *Densovirus* genus (27). NS transcription and splicing followed the same strategy as that previously described for GmDNV (28). A large unspliced transcript (nt 192 to 2596) was found to code for NS3 (first AUG in ORF-3), starting at nt 225. The NS3 coding sequence was removed in the spliced form in roughly half of these transcripts, with an intron from 221-AG/GT to 853-CAG, resulting in a 1,770-nt transcript that was able to code for NS1, starting at the new first codon (856-AUG), and NS2, starting at 875-AUG, by a leaky scanning mechanism due to the poor environment of the NS1 AUG codon (cagAUGa) and the strong environment of the downstream NS2 initiation codon (AcgAUGG). These two maps confirmed the sizes of the mRNAs observed by Northern blotting and indicated that this virus expressed NS1-3 in a fashion identical to that for other *Densovirus* members.

The single band of VP transcripts observed by Northern blotting could actually represent different forms of mRNA with similar intron sizes. The unspliced form, starting at nt 5235, would be 2,672 nt [plus the poly(A) tail] long. First, we determined whether ORF-A and ORF-B could be connected in

frame by splicing, using RT-PCR with primers Adsp and AdlgF (Table 1). For all parvoviruses studied in this respect, VP1 is an N-terminally extended form of VP2, and the position of the Adsp primer was therefore chosen about 25 nt downstream of the VP2 start codon in ORF-A, with AdlgF located at the start of ORF-B. Two alternative introns, with sizes of 131 and 356 nt, were found with the same splice acceptor at codon 4405-CAG (Fig. 1D and 3A and B). Both introns failed to yield an in-frame coding sequence with ORF-A. The stop codon in the spliced ORF-B overlapped the start codon for VP2 (ugUAAUGa) (Fig. 1D). In some systems, e.g., influenza B virus (19, 20) and some non-long-terminal-repeat (non-LTR) retrotransposons (12), reinitiation occurs at such stop-start sequences. Expression of the intronless sequence from nt 4546 (before the small intron) to nt 3892 in the baculovirus system via pFastbacHTb, using primers Adsvp and NAdR (Table 1) and cloning using EheI and EcoRI sites, did not yield products larger than the 29 aa expected from the baculovirus/ORF-B construct, arguing against reinitiation. These results gave credence to the previously suggested ribosome frameshift for PfDNV to generate a nested set of N-terminally extended structural proteins, as observed for all parvoviruses studied thus far (33). To test this hypothesis, we made several recombinant baculovirus constructs such that their expression products would be amenable to N-terminal sequencing in the potential frameshift region and they would be of reasonable size for mass spectrometry (Fig. 1E).

Expression of VP1. Figure 1E shows the 43-nt overlap of ORF-B with the N-terminal extension of ORF-A and the positions of primers (after the ORF-B and before the ORF-A stop codon) used to study the potential translational frameshift. The pFastbacHTb vector was used to yield products that could be purified via their N-terminal His tails and cleaved with the tobacco etch virus (TEV) protease, leaving only one codon (Gly) upstream of the insertion at the blunt-end EheI restriction site (Fig. 4A). The PCR products obtained using the forward primers Ad3s, Ad6s, Ads, Adm, and Adl (Table 1) (the distance from the ORF-B stop codon is indicated, in codons, in Fig. 4A) and the reverse primer NAdR, chosen at the beginning of the extension of VP4 to ensure stable prod-

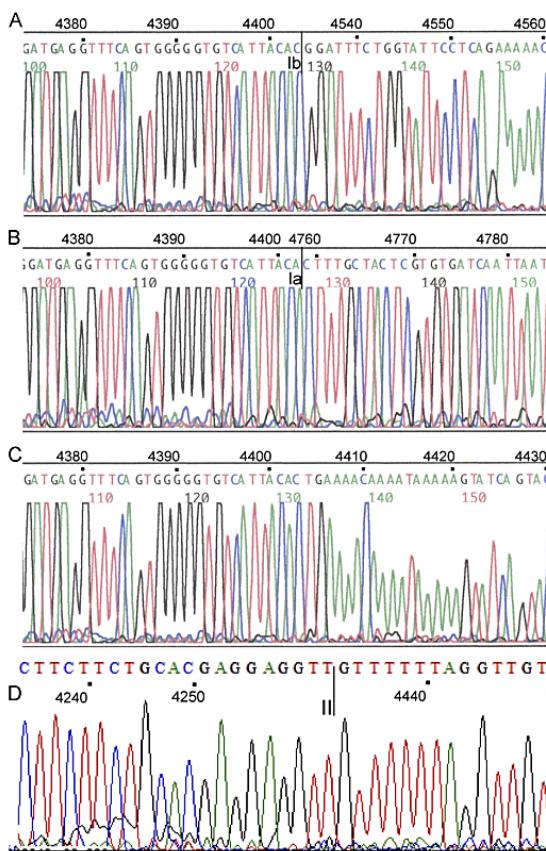


FIG. 3. (A to D) Sequences of mRNAs in the VP ORFs, i.e., removal of the Ib intron in ORF-B (A), removal of the Ia intron in ORF-B (B), unspliced mRNA (C), and the connection of ORF-A with ORF-B by splice II (D). (E) Mascot search results after mass spectrometry for capsid proteins from viruses in natural infections. The proteins were treated with trypsin, which cuts at the C-terminal side of K or R unless the next residue is P. Matched peptides are shown in bold red.

ucts, were cloned into the EheI site, and the clones were selected for orientation by XbaI analysis (the XbaI sites in the MCS and the NAdR primer should be close to each other) and then verified by sequencing.

All constructs (His₆-TEV recognition site-EheI site-insert-XbaI site) yielded proteins that could be purified via their His tails and that had molecular masses indicating that the two frames were connected. The Ad3s and Ads products were treated with TEV protease and were subjected to N-terminal sequencing after the N-terminal His tail/TEV site fragments and the protease were removed by affinity chromatography on Ni-nitrilotriacetic acid (Ni-NTA) columns and analyzed by SDS-PAGE (Fig. 4B). The sequences obtained (Ad3s sequence, GQP?RAEE; and Ads sequence, GAPQQP??QPPaAE), containing an N-terminal G and GA that were introduced via the primers and remained after TEV cleavage, indicated that nt 4435 in ORF-B was connected in frame with nt 4259 in ORF-A. Mass spectrometry analysis of the purified Ads product yielded a mass of 13,652 Da, with masses of 13,732 and 13,812 Da for phosphorylated species, for a predicted mass of 13,649.64 Da for the sequence GAPQQP...GGKRSR (Fig. 4C). These results confirmed the occurrence of a splice between nt 4435 and nt 4259. The predicted mass of VP1 is thus

E Mass spectrometry of capsid proteins

VP1

```

1 MSGVFTDLTL RDFLSLGTNNL PLANTPKLNR RFGRWWNRSR PYDRLPTNEP EPLRETSFOE
61 AAGPEETRIID IAEDEINAGE GAAQAETSFs TGVEETALEL GEAATETTGL LGGATAGSAA
121 AGTAGTLGTV AATAAGGAAL AGIGIGIKKL IDHTSSKGAV LPGTDVFVPGP NFIDPKPARS
181 ETDQIAKEHD LGYEDLLLRA K$QYFTEEDF KTEVYVLLDDE AIHRFSEEQY KSGTWQAFVG
241 KYGLKAKRRL EDVIGGPYP QQPKKQAPRA EEEEVAPATP TADDILDEL PEGNEDDIDF
301 DVDFQLPSTTQ EHHPPTDAA QPGRSTDPP EPVSNFIDST PPDEGDSSSM SAPLTPPITP
361 DKTQDKGKGK RGGGGRPPKS SGGKRSRMA TSLPGAGADI AGEGGGQAI PIPRPITPPH
421 RFLTYGLANV VVIPITRTVN VTIVDNFVIT SLGRLPVDRP FLYMNPNEFA
481 QLPPGSCIECN VRVRLTAFTP RIAFQTNSSN TGLATLNQNQ FILHATGLNI KTQGVDRVPK
541 EFQANEPMVV SSIDELGAIR DENLFEDYVK EFYGDKSVPN HVPRHQFGIP YPLKNYFALV
601 NVQNGTADNP YECLQSHVN EIRCNGPPGE IVEYNKYKQL GLIKKAIAPV YTGVPSAVGG
661 NSTLSCPAGA GNSQYRTADF TMTNNILRG SETFKDTVVS NTEWTIATQ EKSQILHAGA
721 FPSYTPRAQ SLHIGVKFVH ALTtanLDDN LNNAANFTDQ AYFIVEAAC VRLQYPSIRP
781 LFNGPNTIPD NQIYSSSTLQ YTHAVGKSTI MGLQQK

```

VP2

```

1 MTPPLKPHQ P ERDNWEYLNE GQRRYAVEQW QLARVRRGLP IDHPIPEPPR AEEEVAPAT
61 PTADDILPEF PEGNPFMIDF FDVDFLPSSTT QEHHPPTDAA AQPGRSTDPP PEVSNFIDS
121 TPPDEGDSSS MSAPLTPPIT PDKTQDKGKG KRGGGRPPK SGGKRSRMA GTSLPGAGAD
181 IAGEGGGQAI IPIPRPITP HNNFIIIFNKV HRFLTYGLAN VVIPITRTVN DTIVDNFVII
241 TSLGRLPNPF AQLPPGSCIE ECNVRVTAFT PIAFQTNSSN TGLATLNQNQ QFILHATGLN
301 QFILHATGLN IKTQGVDRP KEFQANEPMV VSSIDELGAI RDENLFEDYV KEFYGDKSVP
361 NHVPRHQFGI PYPLKNYPAL VLQNGTADNP PGYECLKSHV NEIRCNCPG IVEYNKYKQL
421 LGLIKKAIPA VYTGVPSAVG GNSTLSCPAG AGNSQYRTAD FTMTNNILRG DSETFKDTVV
481 SNTTIAQ EKSQILHAG AFPSYTPRAQ PSLHIGVKFV HALTTANLDD NLNNANFTDQ
541 QAYFIVEAAC KVRLQYPSIR PLFNGPNTIP DNQIYSSSTL QYTHAVGKST IMGLQQK

```

VP3

```

1 MSAPLTPPIT PDKTQDKGKG KRGGGRPPK SGGKRSRMA GTSLPGAGAD IAGEGGGQAI
61 IPIPRPITP HNNFIIIFNKV HRFLTYGLAN VVIPITRTVN DTIVDNFVII TSLGRLPVDR
121 PFPLYMNPNEF AQLPPGSCIE ECNVRVTAFT PIAFQTNSSN TGLATLNQNQ QFILHATGLN
181 IKTQGVDRP KEFQANEPMV VSSIDELGAI RDENLFEDYV KEFYGDKSVP NHVPRHQFGI
241 PYPLKNYPAL VLQNGTADNP PGYECLKSHV NEIRCNCPG IVEYNKYKQL LGLIKKAIPA
301 VYTGVPSAVG GNSTLSCPAG AGNSQYRTAD FTMTNNILRG DSETFKDTVV SNTTIAQ
361 RGDSETFKDT VVSNTTIAQ TQIEKSQILH AGAFPSYTPR AQPSLHIGVK PVHALTTANL
421 KVRLQYPSIR PLFNGPNTIP DNQIYSSSTL QYTHAVGKST IMGLQQK

```

VP4

```

1 MAGTSLPGAG ADIAGEGGQQ AIPIPPIIT PTHNNFIIFN KVHRFLTYGL ANVVIPITRT
61 VNDVITYVDNF VITSGLRPLV DRFLYMFN EFQPLPVGSC IEECNVRVTA FTPIRAFQTN
121 SSNTGLATN QNQFILHATG LNIKTFQGVDV RPKEFQANEPM MVVSSIDEGL AIRDENLFED
181 YVKEFYGDKS VPNVPRHQF GIPYPLKNYF ALVLQQNGTA DNPGYECLQS HVNEIRCNCPG
241 PGEIVEYNK PQLGLKAI PAVYTGVPSA VGGNSTLSCP AGAGNSQYRT ADFTMTNNIL
301 RGDSETFKDT VVSNTTIAQ TQIEKSQILH AGAFPSYTPR AQPSLHIGVK PVHALTTANL
361 DDNNLNANFT DTDQAYFIVEA ACKVRLQYFS IRPLFNGPNT IPDNQIYSSS TLQYTHAVGK
421 STIMGLQQK

```

88 kDa less than that estimated by SDS-PAGE (Fig. 2A), which may be explained by the phosphorylation observed by mass spectrometry.

Splicing was further investigated by RT-PCR of mRNAs extracted from infected crickets, using 2 sets of primers, namely, AdlgF/NAdR and Ads/AdVPR (Table 1), covering the whole coding sequence of VP1 except for the common VP4 sequence, with estimated products of 1,357 and 1,847 bp without splicing and 1,182 and 1,672 bp after intron II splicing, respectively. The intron II splice was also confirmed by sequencing of the VP1 cDNA (Fig. 3D). As illustrated in Fig. 1D and E, the Ia or Ib intron and the intron II splice were mutually exclusive, i.e., the intron II splice removed the acceptor site for the two intron I splices, whereas the intron I splices removed the donor site for intron II splicing. This expression strategy was further confirmed using recombinant baculovirus constructs. The VP1 sequence from which intron II was removed, and which was thus rendered resistant to ORF-B splicing (Ia and Ib), did not yield VP2 (Fig. 5A). In contrast, constructs with a mutated VP1 initiation codon and a normal VP template yielded a strong VP2 band but also some VP3 and VP4 (Fig. 5A), because type II splicing could remove the VP2 ini-

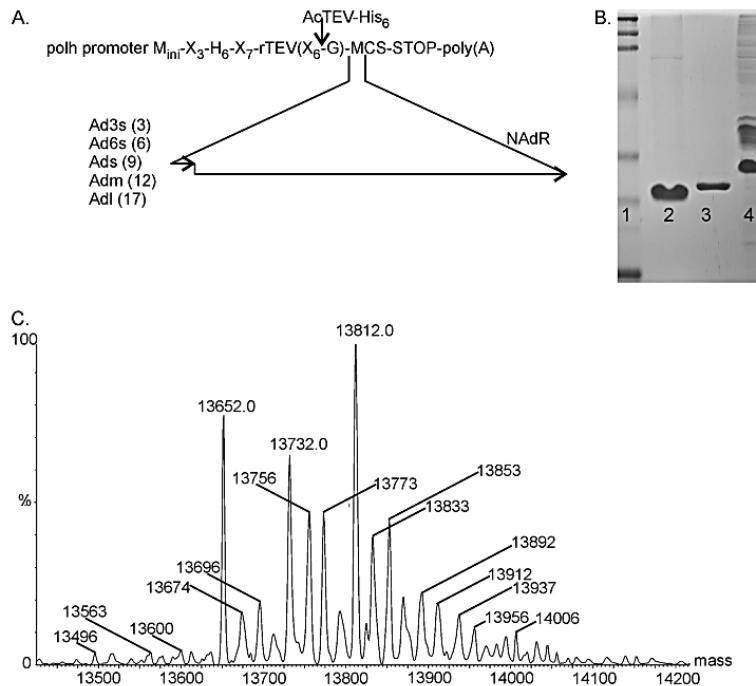


FIG. 4. (A) Construction of various ORF-A and ORF-B constructs with different Ad primers (Ad3s, Ad6s, etc.) linking a short part of ORF-B to a larger part of ORF-A, using the NAdR primer and the pFastbacHTb baculovirus system. Upstream of the insert is a TEV proteolytic site and a histidine tail (H6); N-terminal sequencing of the digested purified protein would reveal the connecting sequences. The numbers in parentheses with the different primers (Ad3s, Ad6s, etc.) denote the distances in codons from the ORF-B stop codon (small arrow). (B) SDS-PAGE analysis of recombinant Ads3 protein generated with Ad3s and NAdR primers, respectively, using the pFastbacHTb baculovirus system. Lane 1, protein markers; lane 2, Ads3 protein after AcTEV protease digestion and purification on a Ni column; lane 3, AcTEV protease (contains a His tag for easy removal after the reaction); lane 4, Ads3 protein prior to cleavage with AcTEV. (C) Mass spectrometry of TEV-treated and purified Ads3 protein, which was predicted from its sequence to have a mass of 13,650.0 Da. A major peak at 13,652 Da verified the predicted sequence. Two other major peaks, at 13,732 Da and 13,812 Da, had an additional 80 and 160 Da, indicating the addition of one and two phosphate groups, respectively. Minor bands had one or more protons replaced by sodium, each adding an additional 22 Da.

itation codon and thus allow downstream initiation. When the template without intron II was used in combination with the mutated VP1 sequence, it yielded, as expected, VP3 and VP4 only. The leaky scanning of the VP3 initiation codon can probably be explained by its weak environment (uccAUGA), in contrast to the strong environment of the VP4 initiation codon (agaAUGG). Therefore, the VP multicistronic cassette yielded 2 sets of structural proteins (Fig. 5B).

Analysis of promoter activity. The promoter elements as well as the poly(A) signals were predicted by the mapping of transcription starts and polyadenylation sites of both NS and VP transcripts (Fig. 1B to D). To assess and compare their functionality, promoter regions (including the start of transcription) were amplified by PCR and cloned upstream of the luciferase gene in the pGL3-basic system. Their relative activities were determined by luciferase assays in independent duplicates at either 40 or 60 h posttransfection. The activity of the NS promoter of AdDNV was very significant in Ld652 cells from gypsy moths but was lower than that of the NS promoter of MIDNV, a lepidopteran densovirus (Fig. 2C). However, the VP1 promoter activity was significantly lower, suggesting the need for *trans*-activation, the absence of a critical factor reacting with the non-ITR region of the VP1 promoter, or differ-

ences in transcription factors between the cricket and gypsy moth systems.

Mass spectrometry of AdDNV capsid proteins. AdDNV was purified and the proteins separated by SDS-PAGE and subjected to mass spectrometry analysis in order to confirm the results obtained by analyzing the baculovirus constructs of the viral proteins. The proteins purified from the gel were digested with trypsin, and sequences of the peptides were determined. Analysis of VP1 and VP2 confirmed the results obtained with the baculovirus expression experiments with respect to the in-frame linking of ORF-B and ORF-A. The peptide sequences obtained covered 33% of VP1, 26% of VP2, 50% of VP3, and 42% of VP4 (Fig. 3E). One outlier peptide identified for VP3 was found in VP2, but with an ion score of 4, this was considered background.

Splicing of the Ia and Ib donor sites with the intron II acceptor site would theoretically also be possible and would yield additional products of 783 and 708 amino acids, close to the 816 amino acids observed for VP1. However, only 4 structural proteins were observed, and mass spectrometry demonstrated the presence of two VP1-specific peptides located in the introns of these potential supplementary products, of 708

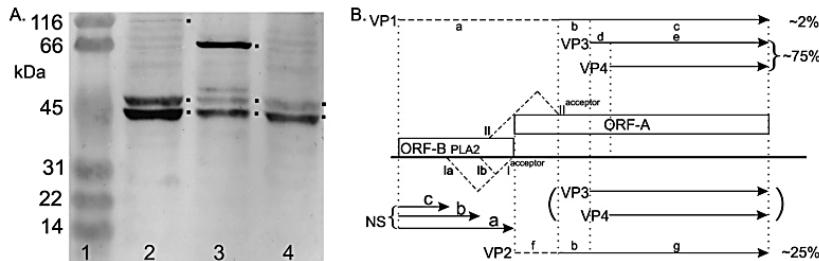


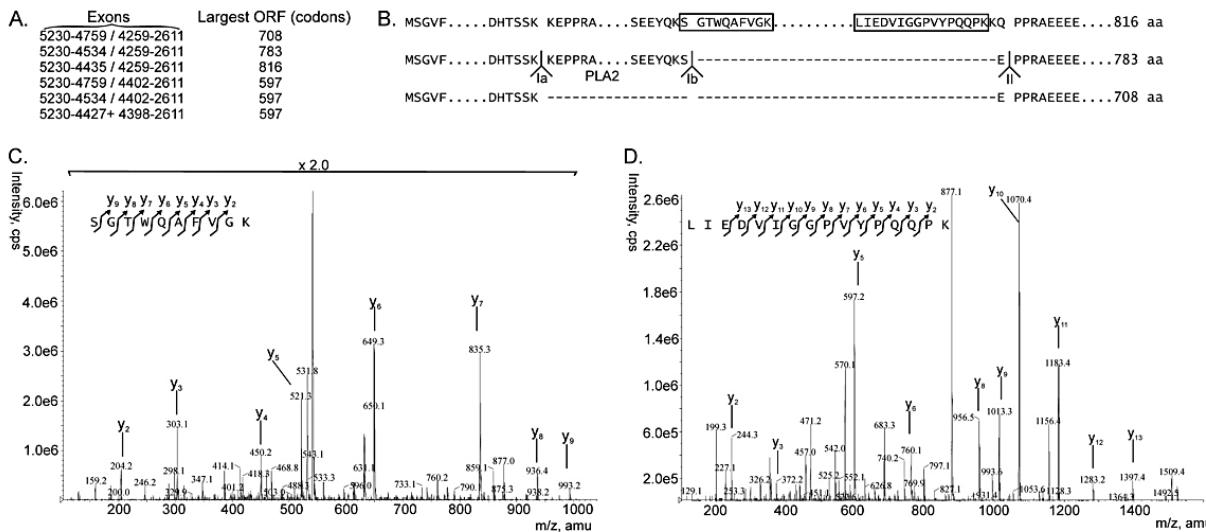
FIG. 5. (A) Western blot of pFastbac VP constructs. Lane 1, prestained SDS-PAGE protein standards (Bio-Rad). Lane 2, expression products from the ORF-A-ORF-B construct from which intron II was deleted. Strong bands were observed for VP3 and VP4, with a weak band for VP1 (all indicated with dots), in addition to some spurious bands. Lane 3, expression products from the ORF-A-ORF-B construct with a mutated VP1 initiation codon. A strong band was observed for VP2, but relatively strong bands were also seen for VP3 and VP4, since some type II splicing occurred that removed the VP2 initiation codon. A weak band was also observed just above VP3. Lane 4, ORF-A-ORF-B construct with a mutated VP1 initiation codon and with intron II deleted, yielding only VP3 (weak initiation codon) and VP4. This lane also contained some of the spurious, nonspecific bands seen in lane 2. (B) Schematic representation of expression products of the VP cassette. Depending on the splice (Ia and Ib versus II), different sets of structural proteins (VP1, VP3, and VP4 versus VP2) were generated. VP3 and VP4 only were expressed in specific constructs (see panel A, lane 4). In addition, expression products solely from ORF-B (NSA, NSB, and NSC) could be expected if initiation at the VP1 codon was combined with Ia or Ib or if no splicing occurred. The predicted pIs of the common protein and N-terminal extensions differed considerably (a, 4.90; b, 3.55; c, 8.58; d, 11.60; e, 6.21; f, 6.72) and may have been a factor in the difference in observed and predicted masses of the capsid proteins (see Fig. 2A). The dashed N-terminal extensions of VP1 and VP2 denote their unique sequences.

and 783 amino acids (Fig. 6), confirming that the 816-aa species corresponded to VP1.

DISCUSSION

The 1977 isolate of AdDNV was cloned and its expression strategy analyzed. Additional AdDNV isolates from Europe,

isolated in 2004, 2006, 2007, and 2009, and from North America, isolated in 1988 (Tennessee) and 2009 (Quebec, Alberta, British Columbia, and Washington State), were amplified by PCR targeting the region between the ITRs and then sequenced (reported elsewhere [24]). All 2009 North American isolates had identical sequences, suggesting a common source, and differed from the 1977, 2004, and 2006 isolates by 49, 18,



and 16 nt substitutions, respectively. The genome organizations of these isolates were identical.

The sequences of the AdDNV ORFs were compared, using nBLAST, with those of other densoviruses, in particular with those of viruses such as PfDNV, PcDNV, BgDNV, MpDNV, and *Dysaphis plantaginea* densovirus (DplDNV), which also have split VP ORFs (see Fig. S3 in the supplemental material). Surprisingly, the highest identities by far for the AdDNV NS1 and ORF-A proteins were found for proteins of PcDNV from *Planococcus citri* (a citrus mealybug belonging to the Pseudococcidae family of the Hemiptera insect order, whereas *Acheta domesticus* belongs to the Gryllidae family of the insect order Orthoptera).

The VP transcript was found to start 23 nt upstream of the 3'-ITR, at nt 5467, and the starts of both NS transcripts were at nt 573, 23 nt downstream of the 5'-ITR. This suggested that many promoter elements would be located within the ITRs and would be identical for the VP and NS promoters. The sequence context of both starts corresponded reasonably well with the consensus sequence for Inr boxes (TCAGTG); however, the promoter activity in lepidopteran cells differed considerably (Fig. 2C). The region from the 5'-untranslated region (5'-UTR) in the VP mRNA to the putative VP1 AUG was only 5 nt long, whereas for the two NS transcripts, the 5'-UTRs were 32 (1.8-kb transcript) and 30 (2.5-kb transcript) nt long.

The expression strategy of the NS cassette is identical to that for the other members of the *Densovirus* genus. In the unspliced transcript (Fig. 1A), NS3 is translated, whereas in the spliced form the ORF for NS3 is removed and translation starts at the weak initiation codon of NS1 or, due to leaky scanning, at the coding sequence for NS2, 19 nt downstream.

In contrast to this conserved strategy, VP expression has unique features that so far have not been observed for vertebrate parvoviruses and densoviruses, which all have a perfect nested set of N-terminally extended structural proteins. AdDNV displays split VP ORFs, and its two largest structural proteins have different extensions to which no roles have yet been assigned. PfDNV (33), PcDNV (25), BgDNV (9, 17), and MpDNV (31, 32), which all have a split VP ORF, with the smaller ORF encoding the PLA2 motif, probably have similar expression strategies (27). For PfDNV, a large number of donor and acceptor splicing sites have also been identified in the VP ORFs. Splicing of the sd3 and sa3 splicing sites in cDNA11 of PfDNV could also yield a large VP1 protein. Like the case for AdDNV, many of the splicing combinations could be inconsequential for generation of the structural proteins. For BgDNV (Dmitry V. Mukha, personal communication), the larger of the two VP ORFs also codes for VP2, and splicing that is slightly different from the strategy in AdDNV also brings the two ORFs in frame and codes for VP1.

Furthermore, it was expected that initiation at the VP1 start codon, which was leaky due to the short 5'-UTR, did not depend on the presence of a downstream splice and that the various spliced and unspliced forms of mRNA were equally probable to be translated. This would lead to premature termination of translation in the case of unspliced mRNA (Fig. 3C), yielding a 268-aa ORF-B product (NSa) that is not involved in capsid formation (Fig. 1A). The 2 ORF-B introns would be responsible for 2 additional, minor nonstructural proteins, of 233 (NSb) and 158 (NSc) aa. Splicing of the intron

II acceptor site with the Ia and Ib donor sites was not observed (Fig. 6).

The genome organization and expression strategy of AdDNV place it in a separate genus from *Densovirus*, and probably from *Pefudensovirus*. Unfortunately, a definitive description of the VP expression of PfDNV is still lacking to resolve whether these viruses should be coclassified. The NS cassette structure of AdDNV, PfDNV, BgDNV, and PcDNV is identical to that of members of the *Densovirus* genus but different from that of the CpDNV ambisense densovirus (1), which has been proposed to be classified in a new genus, *Cupidensovirus* (26). However, these viruses with split VP ORFs are unique among all ambisense densoviruses studied so far in that their ITRs form perfect hairpins of about 150 to 200 nt and they have genomes of about 5,450 nt (27). The low sequence identity among the AdDNV-like viruses may not be evocative of a need to classify them into diverse genera.

ACKNOWLEDGMENTS

P.T. acknowledges support from the Natural Sciences and Engineering Research Council of Canada.

We thank the professionals in the cricket-rearing industry for their work.

REFERENCES

- Baquerizo-Audiot, E., et al. 2009. Structure and expression strategy of the genome of *Culex pipiens* densovirus, a mosquito densovirus with an ambisense organization. *J. Virol.* **83**:6863–6873.
- Bergoin, M., and P. Tijssen. 2010. Densoviruses: a highly diverse group of arthropod parvoviruses, p. 57–90. In N. Asgari and K. N. Johnson (ed.), *Insect virology*. Horizon Scientific Press, Norwich, United Kingdom.
- Bergoin, M., and P. Tijssen. 2008. Parvoviruses of arthropods, p. 76–85. In B. Mahy and M. van Regenmortel (ed.), *Encyclopedia of virology*, 2nd ed., vol. 4. Elsevier, Oxford, United Kingdom.
- Farr, G. A., L. G. Zhang, and P. Tattersall. 2005. Parvoviral virions deploy a capsid-tethered lipolytic enzyme to breach the endosomal membrane during cell entry. *Proc. Natl. Acad. Sci. U. S. A.* **102**:17148–17153.
- Fédière, G. 2000. Epidemiology and pathology of Densovirinae, p. 1–11. In S. Faisst and J. Rommelaere (ed.), *Parvoviruses. From molecular biology to pathology and therapeutic uses*. Karger, Basel, Switzerland.
- Fédière, G., Y. Li, Z. Zadori, J. Szélei, and P. Tijssen. 2002. Genome organization of Caspahalia extranea densovirus, a new iteravirus. *Virology* **292**:299–308.
- Fernandes, S., and P. Tijssen. 2009. Seamless cloning and domain swapping of synthetic and complex DNA. *Anal. Biochem.* **385**:171–173.
- Hellman, U., C. Wernstedt, J. Gonez, and C. H. Heldin. 1995. Improvement of an “in-gel” digestion procedure for the micropreparation of internal protein fragments for amino acid sequencing. *Anal. Biochem.* **224**:451–455.
- Kapelinskaya, T. V., E. U. Martynova, A. L. Korolev, C. Schal, and D. V. Mukha. 2008. Transcription of the German cockroach densovirus BgDNV genome: alternative processing of viral RNAs. *Dokl. Biochem. Biophys.* **421**:176–180.
- Kaufmann, B., et al. 2010. Structure of *Penaeus stylostris* densovirus, a shrimp pathogen. *J. Virol.* **84**:11289–11296.
- Kaufmann, B., et al. 2011. Structure of *Bombyx mori* densovirus 1, a silk-worm pathogen. *J. Virol.* **85**:4691–4697.
- Kojima, K. K., T. Matsumoto, and H. Fujiwara. 2005. Eukaryotic translational coupling in UAAUG stop-start codons for the bicistronic RNA translation of the non-long terminal repeat retrotransposon SART1. *Mol. Cell. Biol.* **25**:7675–7686.
- Laemmli, U. K. 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**:680–685.
- Luckow, V. A., S. C. Lee, G. F. Barry, and P. O. Olins. 1993. Efficient generation of infectious recombinant baculoviruses by site-specific transposon-mediated insertion of foreign genes into a baculovirus genome propagated in *Escherichia coli*. *J. Virol.* **67**:4566–4579.
- Matsudaira, P. 1987. Sequence from picomole quantities of proteins electrophoresed onto polyvinylidene difluoride membranes. *J. Biol. Chem.* **262**:10035–10038.
- Meynardier, G., et al. 1977. Virus de type densoviroïde chez les orthoptères. *Ann. Soc. Entomol. Fr.* **13**:487–493.
- Mukha, D. V., A. G. Chumachenko, M. J. Dykstra, T. J. Kurtti, and C. Schal. 2006. Characterization of a new densovirus infecting the German cockroach, *Blattella germanica*. *J. Gen. Virol.* **87**:1567–1575.

18. Mukha, D. V., and K. Schal. 2003. A densovirus of German cockroach *Blatella germanica*: detection, nucleotide sequence and genome organization. *Mol. Biol. (Moscow)* **37**:607–618.
19. Powell, M. L., T. D. Brown, and I. Brierley. 2008. Translational termination-re-initiation in viral systems. *Biochem. Soc. Trans.* **36**:717–722.
20. Powell, M. L., S. Napthine, R. J. Jackson, I. Brierley, and T. D. Brown. 2008. Characterization of the termination-reinitiation strategy employed in the expression of influenza B virus BM2 protein. *RNA* **14**:2394–2406.
21. Simpson, A. A., P. R. Chipman, T. S. Baker, P. Tijssen, and M. G. Rossmann. 1998. The structure of an insect parvovirus (*Galleria mellonella* densovirus) at 3.7 Å resolution. *Structure* **6**:1355–1367.
22. Styer, E. L., and J. J. Hamm. 1991. Report of a densovirus in a commercial cricket operation in the Southeastern United-States. *J. Invertebr. Pathol.* **58**:283–285.
23. Sukhumsirichart, W., P. Attasart, V. Boonsaeng, and S. Panyim. 2006. Complete nucleotide sequence and genomic organization of hepatopancreatic parvovirus (HPV) of *Penaeus monodon*. *Virology* **346**:266–277.
24. Szeli, J., et al. 2011. Susceptibility of North-American and European crickets to *Acheta domesticus* densovirus (AdDNV) and associated epizootics. *J. Invertebr. Pathol.* **106**:394–399.
25. Thao, M. L., S. Wineriter, G. Buckingham, and P. Baumann. 2001. Genetic characterization of a putative densovirus from the mealybug *Planococcus citri*. *Curr. Microbiol.* **43**:457–458.
26. Tijssen, P., et al. Parvoviridae. In A. M. Q. King, M. J. Adams, E. Carstens, and E. J. Lefkowitz (ed.), *Virus taxonomy: classification and nomenclature of viruses: ninth report of the International Committee on Taxonomy of Viruses*, in press. Elsevier, San Diego, CA.
27. Tijssen, P., et al. 2006. Evolution of densoviruses, p. 55–68. In J. R. Kerr et al. (ed.), *Parvoviruses*. Hodder Arnold, London, United Kingdom.
28. Tijssen, P., et al. 2003. Organization and expression strategy of the ambisense genome of densonucleosis virus of *Galleria mellonella*. *J. Virol.* **77**:10357–10365.
29. Tijssen, P., T. Tijssen-van der Slikke, and E. Kurstak. 1977. Biochemical, biophysical, and biological properties of densonucleosis virus (paravavirus). II. Two types of infectious virions. *J. Virol.* **21**:225–231.
30. Towbin, H., T. Staehelin, and J. Gordon. 1979. Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: procedure and some applications. *Proc. Natl. Acad. Sci. U. S. A.* **76**:4350–4354.
31. van Munster, M., et al. 2003. Characterization of a new densovirus infecting the green peach aphid *Myzus persicae*. *J. Invertebr. Pathol.* **84**:6–14.
32. van Munster, M., et al. 2003. A new virus infecting *Myzus persicae* has a genome organization similar to the species of the genus Densovirus. *J. Gen. Virol.* **84**:165–172.
33. Yamagishi, J., Y. Hu, J. Zheng, and H. Bando. 1999. Genome organization and mRNA structure of *Periplaneta fuliginosa* densovirus imply alternative splicing involvement in viral gene expression. *Arch. Virol.* **144**:2111–2124.
34. Zadori, Z., et al. 2001. A viral phospholipase A2 is required for parvovirus infectivity. *Dev. Cell* **1**:291–302.

The Structure and Host Entry of an Invertebrate Parvovirus

Geng Meng,^a Xinzhen Zhang,^a Pavel Plevka,^{a*} Qian Yu,^b Peter Tijssen,^b Michael G. Rossmann^a

Department of Biological Sciences, Purdue University, West Lafayette, Indiana, USA^a; INRS-Institut Armand-Frappier, Université du Québec, Laval, Québec, Canada^b

The 3.5-Å resolution X-ray crystal structure of mature cricket parvovirus (*Acheta domesticus* densovirus [*AdDNV*]) has been determined. Structural comparisons show that vertebrate and invertebrate parvoviruses have evolved independently, although there are common structural features among all parvovirus capsid proteins. It was shown that raising the temperature of the *AdDNV* particles caused a loss of their genomes. The structure of these emptied particles was determined by cryo-electron microscopy to 5.5-Å resolution, and the capsid structure was found to be the same as that for the full, mature virus except for the absence of the three ordered nucleotides observed in the crystal structure. The viral protein 1 (VP1) amino termini could be externalized without significant damage to the capsid. *In vitro*, this externalization of the VP1 amino termini is accompanied by the release of the viral genome.

Parvoviruses are small (~250- to 300-Å-diameter), single-stranded DNA (ssDNA), icosahedral (T=1), nonenveloped viruses whose genomes are approximately 5 kb long (1). The *Parvoviridae* family has been subdivided into viruses that infect vertebrates (*Parvovirinae*) and those that infect invertebrates (*Densovirinae*) (2). Parvoviruses replicate in dividing cells such as in tissues from insect larvae and fetuses. Densovirus are highly pathogenic, and those that use insect hosts usually kill 90% of the larvae within a few days (2). Densovirus pose a threat to commercial invertebrates such as shrimp (3), silkworms (4), and crickets (5, 6). Some highly pathogenic densovirus are potential selective pesticides for vectors that transmit mosquito-borne diseases (7). *Parvovirinae* generally have three types of proteins (VP1, VP2, and VP3) in their capsids (8), whereas *Densovirinae* generally have four types of proteins (VP1 to VP4) in their capsids (2). In densovirus there are 200 additional amino acids in VP1 at the N terminus. These different proteins result from different initiation sites for translation of the capsid gene and from posttranslational modification of their N termini (8). Generally, each of the 60 subunits within a capsid has the same amino acid sequence and is structurally the same, except that the different proteins start at different amino acids. The VP2s of some densovirus are unique among VP2s of parvoviruses since they are not completely contained within corresponding VP1s (Fig. 1A).

Parvoviruses enter cells by dynamin-dependent receptor-mediated endocytosis and escape the endosome by the phospholipase (PLA2) activity within the amino-terminal domain of VP1 (9–13). Although there is often less than 5% amino acid identity among the structural proteins of parvoviruses, the sequence of the PLA2 N-terminal domain of VP1 has more than 30% amino acid identity (Fig. 1A and B). The PLA2 domain is not exposed in assembled, full parvoviruses such as minute virus of mice (MVM) (13) and human parvovirus B19 (14), and it therefore has to be exposed during endocytosis (9, 11, 13–15). However, the mechanism by which the VP1 amino-terminal PLA2 domain is exposed has not been elucidated in detail (16).

The structures of six autonomous vertebrate parvoviruses (canine parvovirus [CPV] [17], feline parvovirus [FPV]) [18], porcine parvovirus [PPV] [19], MVM [20], H-1 parvovirus [H-1PV] [21], and human parvovirus B19 [22]) and three invertebrate parvoviruses (*Galleria mellonella* densovirus [*GmDNV*] [23], *Bombyx mori* densovirus [*BmDNV*] [24], and *Penaeus stylorostis*

densovirus [*PstDNV*] [25]) have been determined (Table 1). Furthermore, extensive studies have been made of the human adenovirus-associated dependoviruses (26, 27). The structures of these parvoviruses consist of 60 structurally equivalent capsid proteins assembled with icosahedral symmetry. Each capsid protein has a “jelly roll” fold, a motif that is common to many viruses, including the nonenveloped RNA picornaviruses (28) and small RNA plant viruses (29) as well as larger double-stranded DNA (dsDNA) adenoviruses (30), the enveloped bacteriophage PRD1 (31), the fungal virus *Paramecium bursaria* chlorella virus 1 (PBCV-1) (32), vaccinia virus (33), and probably also mimivirus (34). The jelly roll fold is a β-barrel consisting of two opposed antiparallel β-sheets with adjacent-strand BIDG and CHEF, where the strands along the polypeptide chain are named A and B to H. The interior of the barrel is exceedingly hydrophobic.

Parvoviruses have a channel along the 5-fold axes formed by five symmetry-related DE loops (the “DE” loop is between the β-strands D and E). Residues lining the channel are mostly hydrophobic and guide the externalization of a conserved glycine-rich sequence near the amino ends of the VPs (35–37). The loops connecting the β-strands of the jelly roll fold are usually exceptionally large in parvoviruses compared with the loops in picornaviruses (28, 38) and form the exterior of the virus and intersubunit contacts (Fig. 1C). These loops are more variable in sequence than the core jelly roll structure.

Here, we describe the crystal structure of mature virions of cricket parvovirus (*Acheta domesticus* densovirus [*AdDNV*]) at 3.5-Å resolution and the cryo-electron microscopic (cryoEM) structure of the emptied virus at 5.5-Å resolution. We also report on the externalization of the VP1 N-terminal region and subsequent genome release by an increase in temperature.

Received 3 July 2013 Accepted 3 September 2013

Published ahead of print 11 September 2013

Address correspondence to Michael G. Rossmann, mr@purdue.edu.

* Present address: Pavel Plevka, CEITEC, Masaryk University, Brno, Czech Republic.

Copyright © 2013, American Society for Microbiology. All Rights Reserved.

doi:10.1128/JVI.01822-13

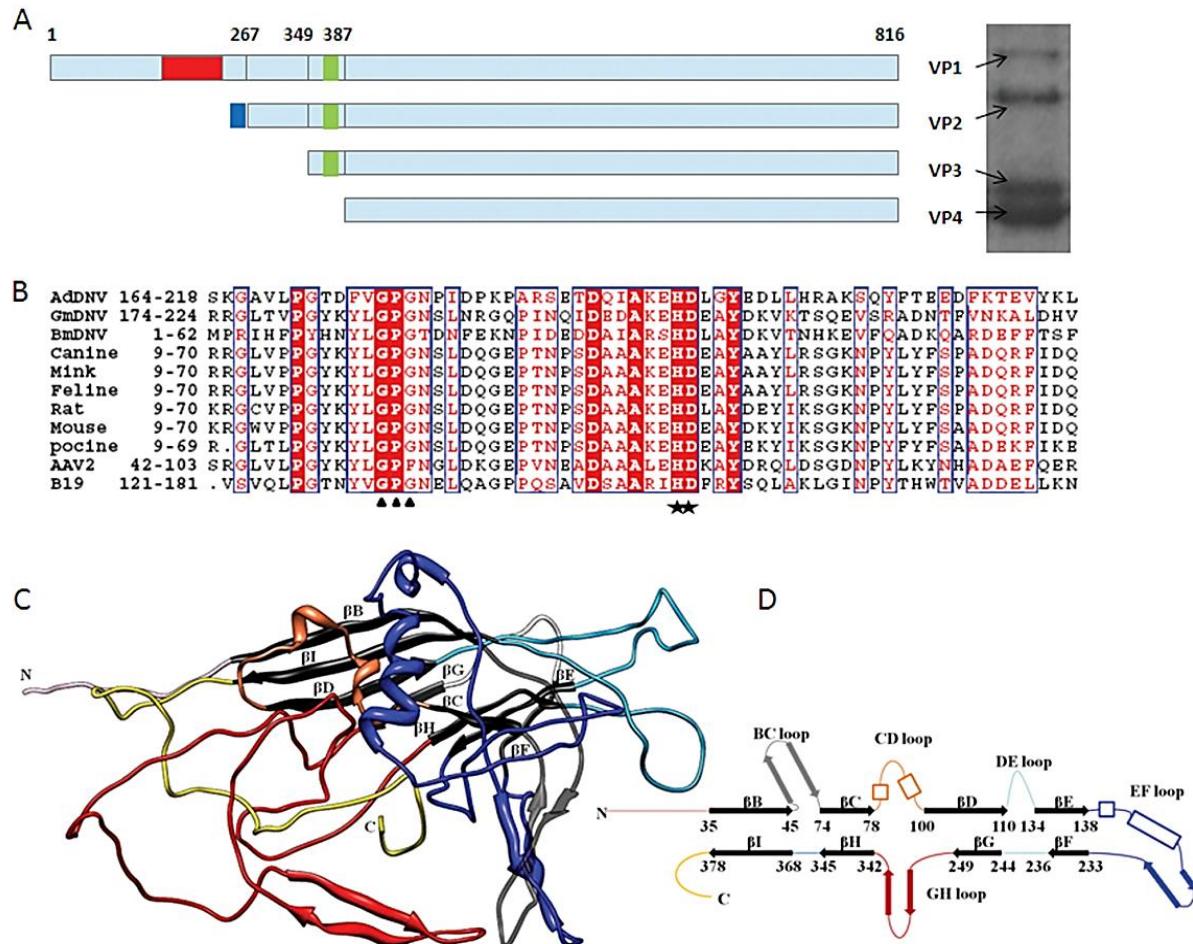


FIG 1 Structure of the AdDNV capsid protein. (A) Left, capsid protein organization. Red area, PLA2 domain; green area, glycine-rich region; blue area, extra sequence at the N termini of VP2. Right, SDS-polyacrylamide gel of the AdDNV full particles. (B) Alignment of VP1 sequences in VP1s of various parvoviruses, including AdDNV, GmDNV, BmDNV, CPV, mink parvovirus, FDV, rat parvovirus, mouse parvovirus, PPV, AAV2, and B19. The stars indicate the His-Asp catalytic site, and the triangles locate the Ca^{2+} -binding site. (C) Three-dimensional structure of AdDNV capsid protein, showing the core jelly roll in black. The surface loops connecting the strands of the core jelly roll are colored as follows: BC loop, gray; CD loop, orange; DE loop, sky blue; EF loop, dark blue; and GH loop, red. (D) Diagrammatic representation of the capsid protein structure (heavy lines represent β -strands) (color coding is the same as in panel C).

MATERIALS AND METHODS

Virus purification and preparation of the emptied virus particles. The original virus was isolated from infected crickets (5). Further purification was achieved using CsCl equilibrium density gradient centrifugation. Of the two bands with different densities, the lower band contained the full particles and represented 99% of all the particles. The upper band contained empty particles assembled mainly from VP4. The two bands were separately transferred into Tris-buffered saline (TBS) (10 mM Tris-Cl, 100 mM NaCl, 1 mM CaCl, and 1 mM MgCl at pH 7.5) for further usage. Aliquots of the full virus particles were incubated at 26, 37, 45, 55, 65, 75, and 100°C for 1 h. The heat-treated emptied particles were frozen on holey carbon (Quantifoil) EM grids and checked by cryoEM. The numbers of full, empty, and broken particles were counted by eye (Fig. 2) and averaged over three holes on two different EM grids. Each hole had roughly 100 particles.

Determination of the crystal structure of the full virus particle. Crystals of the full particles were obtained by hanging-drop vapor diffu-

sion in the presence of 20% polyethylene glycol (PEG) 400 and 100 mM MgCl_2 at 16°C. Further optimization of the crystallization conditions produced crystals of up to 0.5 mm in length. Crystals were soaked for at least 20 min in the presence of 20% glycerol cryoprotectant prior to freezing.

X-ray diffraction data were collected at 100 K at the Advanced Photon Source (APS) beamline 23ID (Table 2). Diffraction data from about 20 crystals were indexed and merged, using the HKL2000 computer program (39) to generate the final 3.5-Å resolution data set. The space group was $P_{4_2}2_1$ with $a = 412.67$ Å and $c = 278.80$ Å. The Matthews coefficient was 3.64 Å³/Da, assuming half a virus particle per crystallographic asymmetric unit. Thus, the virus was located on a crystallographic 2-fold axis. A self-rotation function, calculated with the GLRF program (40) using 8- to 3.5-Å resolution data, gave the accurate orientation of the particle about the crystallographic 2-fold axis. This showed that one of the icosahedral 2-fold axes of the virus was roughly parallel to the 4_2 crystallographic axis, with a 1.6° rotation away from being exactly parallel. As a consequence, the position of the particle along the crystallographic 2-fold axis could be

TABLE 1 Structural studies of autonomous parvoviruses

Virus	Description of particle	Structural protein(s) in particles	Resolution (Å)	Icosahedral ordered genome structure (bp)	PDB code (reference)
Vertebrate parvoviruses					
Canine parvovirus	Full virus	VP1, VP2, VP3	2.9	11	4DPV (17)
	Empty particle	VP1, VP2	3.0	None	2CAS (56)
Feline parvovirus	Empty particle	VP3	3.3	None	1FPV (18)
Porcine parvovirus	Virus-like particle	VP2	3.5	None	1K3V (19)
Human parvovirus B19	Virus-like particle	VP2	3.5	None	1S58 (22)
Minute virus of mice	Full virus	VP1, VP2, VP3	3.5	11	1MVM (20)
Rat H-1 parvovirus	Full virus	VP1, VP2, VP3	2.7	10	4GOR (21)
	Empty particle	Unknown	3.2	None	4GBT
Invertebrate parvoviruses					
GmDNV	Full virus	VP1, VP2, VP3, VP4	3.7	None	1DNV (23)
BmDNV	Virus-like particle	VP3	3.1	None	3POS (24)
PstDNV	Virus-like particle	VP4	2.5	None	3N7X (25)
AdDNV	Full virus	VP1, VP2, VP3, VP4	3.5	3	
	Induced emptied particle	VP1, VP2, VP3, VP4	5.5	None	

determined from the big Patterson peak generated by the large number of parallel equal-length vectors.

The structure was determined using the molecular replacement method (41) with the structure of *GmDNV* (Protein Data Bank [PDB] code 1DNV) (23) as the initial phasing model to 15-Å resolution. The phases were then extended to 3.5-Å resolution in steps of one reciprocal lattice interval (1/c) at a time. Three cycles of 30-fold noncrystallographic symmetry (NCS) averaging and solvent flattening were performed for each extension step. The averaging and extension processes were performed using the program AVE in the Uppsala Software Factory (42) and

FFT, SFALL in CCP4 programs (43). The final overall correlation coefficient between the observed structure amplitudes and the calculated structure factors corresponding to the final averaged and solvent-flattened map was 0.866. The atomic model was built into the 3.5-Å resolution map using COOT (44). The model coordinates were refined with the CNS program (45) while applying NCS constraints and reasonable model restraints, including group temperature factor refinement. No attempt was made to identify water molecules, as the data extended to only 3.5-Å resolution. The structures of the three ordered nucleotides bound to the inside surface of the capsid (see Results and Discussion) were included in

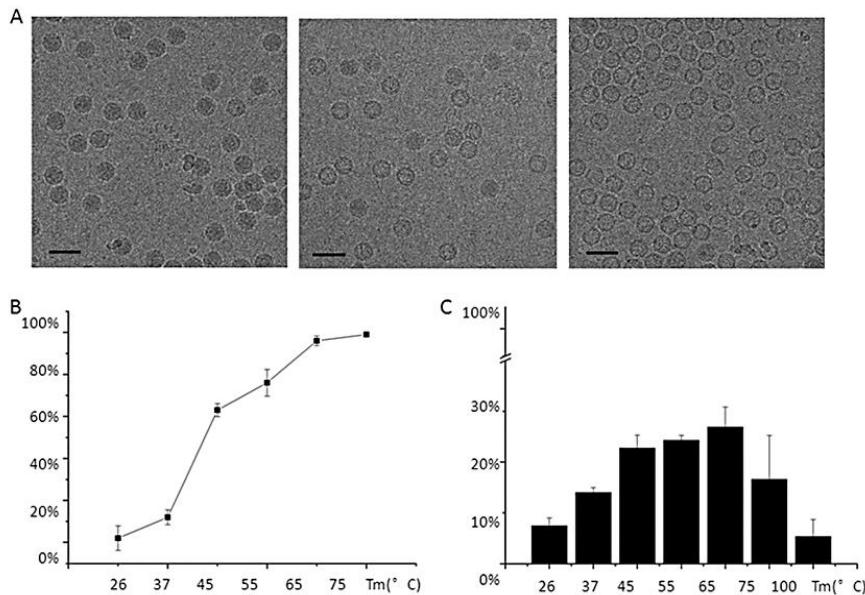


FIG 2 Heat treatment and PLA2 activity of the virus particles. (A) CryoEM micrographs showing the virus after incubation for 1 h at 26°C, 45°C, and 65°C, resulting in only full particles (left), about an equal number of full and emptied particles (middle), and emptied particles (right), respectively. (B) Percentage of emptied particles after incubating full particles at different temperatures. (C) Phospholipase activity as a function of temperature with respect to the PLA2 activity of honey bee PLA2 at 26°C.

TABLE 2 X-ray data collection and structure refinement

Parameter	Value ^a
Data collection parameters	
Wavelength (Å)	1.0715
CCD detector	MAR CCD-325
Exposure time/frame (s)	3
Oscillation angle/frame (°)	0.3
No. of frames collected	300
Data reduction and refinement statistics	
Resolution range (Å)	50–3.5 (3.6–3.5)
Space group	P4 ₂ 2 ₁ 2
No. of frames used	80
Cell parameters	
<i>a</i> , <i>b</i> , <i>c</i> (Å)	412.67, 412.67, 278.80
α, β, γ (°)	90, 90, 90
Mosaicity (°)	0.40
No. of observed reflections	1,518,629
No. of unique reflections	236,217 (9,755)
Redundancy	1.2 (1.1)
% Completeness	56 (46.5)
$\langle I \rangle / \langle \sigma(I) \rangle$	4.0 (1.2)
$R_{\text{sym}} \sum h \sum j I_{hj} - \langle I_h \rangle / \sum \sum I_{hj}$	0.161 (0.565)
Model building and refinement statistics	
Resolution range (Å)	30–3.5 (3.6–3.5)
No. of residues/atoms built	412,3290
Final <i>R</i> factor ^b	0.289
Mean isotropic temp factor (Å ²)	18
RMSD bond length (Å)/bond angle (°)	0.0047/1.258
% Residues in most favored/additionally allowed/generously allowed/disallowed regions of the Ramachandran plot ^c	82.5/16.9/0.3/0.3

^a Values in parentheses refer to the highest-resolution shell.^b No *R*_{free} value was calculated, because the high NCS redundancy interrelates reflections, causing the free reflections to be dependent on all other reflections. As a result, there would be little difference between *R*_{working} and *R*_{free}.^c Percentage of a total of 360 nonglycine, nonproline residues as defined in the program PROCHECK.

the final stages of refinement. After more than 5 cycles of refinement and model rebuilding, the *R* factor had dropped from 34% to 28.9%. In the presence of the 30-fold NCS redundancy, there will be no significant difference between *R*_{free} and *R*_{working}.

Detecting the externalized VP1 N termini and their phospholipase A2 activity. Full particles and the heat-treated emptied particles were digested by trypsin at room temperature. About 30 μl of the particle suspension at a concentration of 5 μg/μl was incubated with trypsin for 1 h at 26°C. The trypsin had a final concentration of 1 μg/μl in the mixture. The samples were then checked for VP1 cleavage using a 15% SDS-polyacrylamide gel (Fig. 3).

The PLA2 activities of the full particles and the heat-treated emptied particles were measured at 26°C by a colorimetric assay (sPLA2 assay kit; Cayman Chemical, Ann Arbor, MI), using the 1,2-dithio analog of diheptanoyl phosphatidylcholine (diheptanoyl thio-PC) as the substrate for PLA2. The absorbance at 405 nm was determined every minute for 30 min. Measurements of the PLA2 activity were normalized relative to the activity of 1 ng bee venom PLA2. The activity of the PLA2 in 1.5 μg virus is equivalent to the PLA2 activity in 1 ng bee venom.

CryoEM and three-dimensional structural reconstruction of emptied particles. The optimal condition for obtaining the largest percentage of emptied, unbroken particles was 55°C for 1 h (Fig. 2). Three microliters of the heat-treated emptied particles at a protein concentration of 5 μg/μl

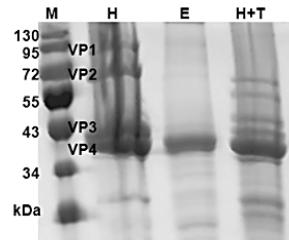


FIG 3 SDS-PAGE of heat-treated AdDNV particles. Lanes: M, protein markers; H, heat-treated emptied AdDNV particles; E, non-heat-treated empty particles among the purified mature AdDNV particles; H+T, further trypsin digestion of the heat-treated emptied AdDNV particles.

was applied to holey grids (Quantifoil) and blotted for 6 s in an FEI Mark 3 Vitrobot chamber at 90% humidity. The grids were then fast-frozen in liquid ethane. Cryo-electron microscopy (CryoEM) images were acquired on an FEI Titan Krios operated at 300 keV. Images were recorded with a 4k × 4k charge-coupled device (CCD) detector. As a control, grids of untreated particles were prepared and viewed in the same way. The assumed magnification of 59,000 was calibrated with respect to a known specimen and was shown to correspond to a pixel separation of 1.51 Å in the image. The electron dose was ~20 e/Å², and the image was defocused by between ~1.6 and 2.6 μm. About 150 cryoEM micrographs, each showing roughly 100 particles, of the emptied particles were recorded. The defocus and the astigmatism of each micrograph were estimated with the EMAN1 fitctf program (46) and further confirmed with the program ctifit. Image processing and three-dimensional reconstruction were performed using the EMAN suite of programs (47). The final reconstruction was computed using ~15,000 particles out of about 17,000 initial boxed images and was found to have 5.5-Å resolution based on the separate structure determinations of two randomly selected independent sets of images using the Fourier shell correlation threshold of 0.143 (Fig. 4) (48).

Sequence alignment of the PLA2 domain and structural comparisons. The sequence of the AdDNV VP1 N-terminal PLA2 domain (GI 326392953) was aligned with the corresponding sequences of adeno-associated virus 2 (AAV2) (GI 110645923), human parvovirus B19 (GI 169212578), CPV (GI 116646110), MVM (GI 332290), rat parvovirus (GI 410443463), mink parvovirus (GI 425696394), PPV (GI 46404508), GmDNV (GI 23334609), and BmDNV (GI 18025360) using Clustal X (49).

The crystal structure of AdDNV was compared with those of other invertebrate densoviruses, i.e., GmDNV (23) and BmDNV (24), as well as

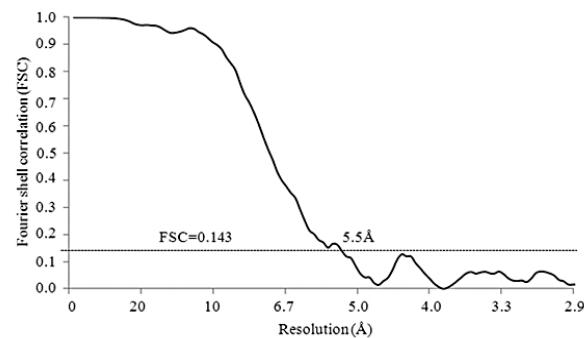


FIG 4 Fourier shell correlation (FSC) based on the independent structure determinations of two randomly selected equally sized sets of images, showing the resolution of the emptied AdDNV particle reconstruction to be ~5.5 Å when the FSC is 0.143.

TABLE 3 Sequence and structural comparisons of *AdDNV* capsid protein with other autonomous parvovirus capsid proteins

Virus	Sequence identity (%)	RMSD (Å) between C _α atoms	No. of aligned C _α atoms	Total no. of C _α atoms
Canine parvovirus	<5	4.8	261	548
Feline parvovirus	<5	4.9	263	534
Porcine parvovirus	<5	4.9	258	542
Human B19	<5	5.1	260	523
Minute virus of mice	<5	5.0	264	549
<i>GmDNV</i>	30.50	2.1	295	415
<i>BmDNV</i>	<5	3.7	331	412
<i>PstDNV</i>	8.30	4.1	224	299

mammalian autonomous parvoviruses CPV (17), PPV (19), FPV (18), MVM (20), and B19 (22) using the HOMOlogy program (50). These structural comparisons do not include the disordered PLA2 domain, whose positions in the virus are random and therefore cannot be observed in the crystal structure.

Accession numbers. The atomic coordinates of the *AdDNV* crystal structures have been deposited with the Protein Data Bank (www.pdb.org) (PDB code 4MGU); the cryo-EM maps of the emptied *AdDNV* particle have been deposited with the Electron Microscopy Data Bank (www.emdatabank.org) (EMDB code EMD-2401).

RESULTS AND DISCUSSION

Crystal structure of the full *AdDNV* particles. The structure of *AdDNV* was determined to 3.5-Å resolution. The position of the core jelly roll relative to the icosahedral symmetry axes was essentially the same in *AdDNV* as in other known parvovirus structures (Table 3 and 4).

The four structural proteins VP1 (88.1 kDa), VP2 (65.3 kDa), VP3 (50.8 kDa), and VP4 (46.9 kDa) are in an approximate 1:11:18:30 proportion in *AdDNV* full particles based on scanning the gel with Kodak Image Station 2000R and analyzing with software Kodak MI (Fig. 1A). The glycine-rich sequence is present in VP1, VP2, and VP3, but is missing in VP4 (Fig. 1A). It may be significant that, compared with vertebrate parvoviruses, there is therefore only one copy of the PLA2 structure per virion. The polypeptide chain of the capsid protein could be traced from residue 23 of VP4 situated at the base of the 5-fold axis channel to residue 418 at the carboxy terminus (Fig. 1C).

The electron density in this channel ($\sigma = 1.5$) of *AdDNV* is weak and discontinuous (Fig. 5C), which is similar to the density in the *GmDNV* 5-fold channel. The glycine-rich motif in *AdDNV* consists of about 17 residues, 8 of which are glycines, whereas in *GmDNV* the same motif has 7 glycines and about 16 residues (Fig. 6). The difference of the sequence length may be partly related to the structure of the channel in the different parvoviruses.

The low density in the 5-fold channel suggests that only several

TABLE 4 Superposition of the *AdDNV* jelly roll core (70 residues) on other invertebrate parvovirus capsid proteins

Virus	RMSD (Å) between C _α atoms
Canine parvovirus	2.0
<i>GmDNV</i>	0.8
<i>BmDNV</i>	1.4
<i>PstDNV</i>	1.6

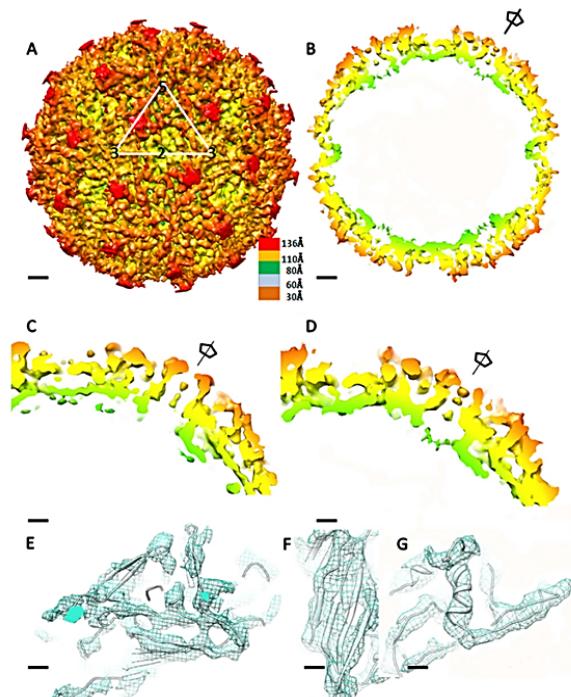


FIG 5 Structure of *AdDNV* emptied particles. (A) CryoEM reconstruction of emptied particles. Surface features with a triangle showing the limits of one icosahedral asymmetric unit are shown. The scale bars represent 2 nm. (B) Center section of the cryoEM reconstruction. The scale bars represent 2 nm. (C) Enlargement of the 5-fold channel density in the X-ray electron density map. The scale bars represent 1 nm. (D) Enlargement of the 5-fold channel density in the cryoEM density map. The scale bars represent 1 nm. (E, F, and G) Fit of the X-ray structure polypeptide backbone into the cryoEM density for the β -sheets of the jelly roll (scale bars represent 5 Å) (E), the BIDG β -sheet (the scale bars represent 3 Å) (F), and the α -helix located in the EF loop (the scale bars represent 3 Å) (G).

of the 12 5-fold channels are occupied, resulting in externalization of the VP amino termini. A similar lack of amino-terminal externalization was observed in *GmDNV*, the only other known structure of a mature DNV. The structures of silkworm and shrimp densoviruses (Table 1) were self-assembled from recombinantly expressed VP3 and VP4 capsid proteins, respectively. Hence, these structures are missing the glycine-rich sequence. As there is only one VP1 per virion, some of the 5-fold channels must be occupied by VP2 or VP3. However, in the vertebrate parvoviruses CPV (17)

AdDNV	GK GKR GCG R PPPKS S GG
<i>GmDNA</i>	G TG SG T SS G GGNT C G
<i>BmDNV</i>	GGGA QV DPR .. TG GQ A GG S GGGMG AGG
canine	GGQPA AVRNERATG SGNS S GG GGGGG S GGVG
mink	GGQPA AVRNERATG SGNS S GG GGGGG S GGVG
feline	GGQPA AVRNERATG SGNS S GG GGGG S GGVG
porcine	G NESGG GG GGGRG AGGVGV VSTG
B19	GAC GGS NPVKS MWSEGAT ..

FIG 6 Sequence comparisons of the glycine-rich regions of *AdDNV* and other autonomous parvovirus, including *GmDNV* (GI 23334609), *BmDNV* (GI 18025360), canine parvovirus (CPV) (GI 116646110), mink enteritis virus (MEV) (GI 425696394), feline parvovirus (FPV) (GI 333476), porcine parvovirus (PPV), and human parvovirus (B19) (GI 169212578).

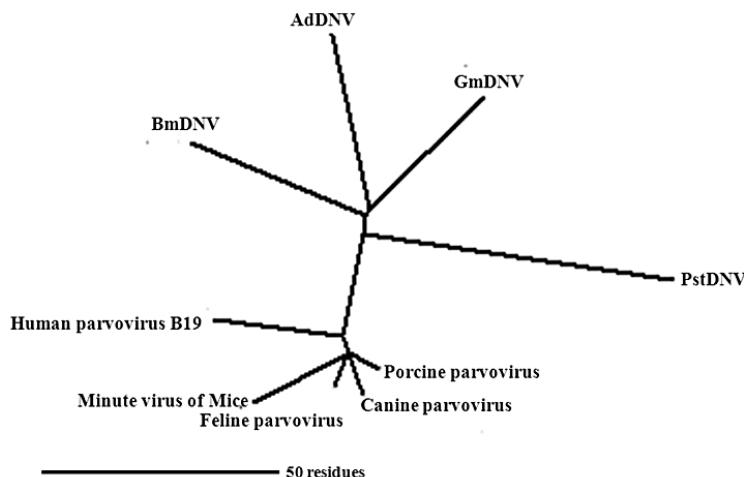


FIG 7 Unrooted phylogenetic tree of different parvoviruses based on the number of inserted residues between β -strands.

and MVM (20), all the 5-fold channels are always fully occupied by the VP amino termini. Furthermore, there is a larger proportion of VP1 subunits per virion in the vertebrate parvoviruses. However, the externalized amino termini are disordered in both the vertebrate and invertebrate parvovirus crystal structures.

Pairwise comparisons of the parvovirus structures were used to determine the number of and root mean square deviation (RMSD) between equivalent Co atoms (Table 3). The number of inserted amino acids between β -strands was used to calculate a phylogenetic tree using the program MEGA (51) (Fig. 7). This shows a closer relationship among the insect densoviruses (*AdDNV*, *GmDNV*, *PstDNV*, and *BmDNV*) than between these viruses and vertebrate viruses. Although the capsid proteins of all parvoviruses have common structural features, the capsid proteins of vertebrate and invertebrate parvoviruses must have evolved independently (Fig. 7).

The A β -strand is folded back to run antiparallel to β -strand B in the BIDG sheet in all known vertebrate parvovirus structures (Fig. 1D). In previously determined densovirus structures, including the *AdDNV* structure reported here, β -strand A is associated with β -strand B in the neighboring, 2-fold related BIDG sheet. Such an exchange of domains between 2-fold-related structures is an example of “domain swapping” (Fig. 1C and D). However, in *AdDNV* the A β -strand contains three proline residues (Pro24, Pro26, and Pro28) and therefore diminishes the H bonding with the B β -strand of the neighboring subunit.

The channel along the 5-fold icosahedral axes of parvoviruses is formed by the DE loop and is between 16 Å and 18 Å in diameter, measured from atom center to atom center. The conformation of the DE loop is variable among both vertebrate and invertebrate parvoviruses (8). In *AdDNV*, as also in all other known parvoviruses, there are several hydrophobic amino acids in the DE loop (from 116 Ala to 133 Gln in *AdDNV*) that interact with the glycine-rich region.

Emptied densovirus particles and phospholipase activity. Most parvoviruses, including densoviruses, assemble *in vivo* both as full infectious particles and as empty particles. However, for *AdDNV* and presumably also for other densoviruses, the small

fraction of particles that are empty in a virus preparation consist of only VP4 (Fig. 3) and are missing the glycine-rich sequence, whereas the dominant infectious virus particles contain all four types of subunits (VP1 to VP4) (Fig. 1A). Therefore, after heat treatment, nearly all the emptied particles that have a full complement of all four VPs must have been full of genome, whereas empty particles containing only VP4 must have been assembled as empty particles. It had been shown that heating parvoviruses to 70°C generated PLA2 activity, suggesting exposure of the VP1 N termini (12, 13). However, it was not clear whether only the VP1 N termini were exposed from intact particles or whether the particles had disassembled. The loss of the genome associated with a presumably transient change in the capsid has some resemblance to the infectious process in picornaviruses (52).

Here we used cryoEM to show that on heating of *AdDNV* for a defined length of time, the number of emptied particles increased with temperature (Fig. 2A and B). When the temperature was increased beyond 65°C there was also an increase of broken particles. Concomitant with the increase of emptied particles, there was also an increase of PLA2 activity (Fig. 2C). Above about 65°C, the virions disintegrated and had reduced PLA2 activity. Unlike the case for full, infectious *AdDNV* particles, the VP1 N termini of the heat-treated emptied particles were sensitive to trypsin digestion, whereas the capsids remained intact as determined with

TABLE 5 Results of a six-dimensional search on fitting of the *AdDNV* capsid protein structure into the 5.5-Å cryoEM density by using the EMfit program (55)^a

Sum ^b	Clash (%) ^c	–Den (%) ^d
40.0	1.7	4.8
24.0	14.6	15.6
21.1	19.4	20.9

^a Three possible fits were found, but the top fit is by far the best.

^b Sumf, mean density height averaged over all atoms, where the maximum density in the electron density map is set to 100.

^c Clash, percentage of atoms in the model that approach closer than 5 Å to icosahedral-related capsid protein molecules.

^d –Den, percentage of atoms in density less than zero density.

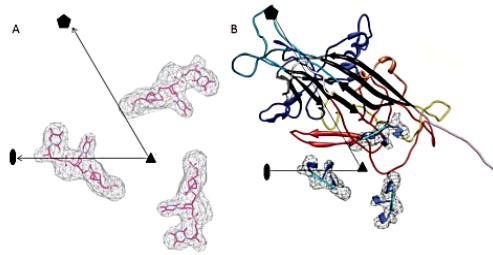


FIG 8 Structure of the three ordered ssDNA bases in AdDNV particles. (A) Difference map between X-ray electron density and cryoEM density, with black lines showing the limits of one icosahedral asymmetric unit. Three bases of the ssDNA were built into the density. (B) Location of the three bases relative to the AdDNV capsid. The colors of the AdDNV capsid polypeptide are as defined for Fig. 1C.

cryoEM. This showed that the heat treatment causes externalization of the N termini while leaving the capsid intact, with the PLA2 domain remaining a part of the particle.

Externalization of the N termini prior to endocytosis abolishes infectivity in parvovirus (53), suggesting that the sequence of events during infection is critical. This could explain why all parvoviruses harbor the N-terminal part of VP1 within the virus particle until they are ready to breach the endosomal membrane. PLA2 requires a Ca^{2+} concentration of greater than 1 mM (9). Such Ca^{2+} concentrations are present in the endosome but not in the cytoplasm (54), further narrowing the viral PLA2 activity to the endosomal membrane.

The 5.5-Å pseudo-atomic-resolution cryoEM structure of emptied AdDNV particles. An initial effort to crystallize heat-treated emptied particles failed to produce crystals, probably because of the externalized VP1 N termini. However, it was possible to obtain a 5.5-Å resolution cryoEM structure from approximately 15,000 images of heat-treated emptied particles (Fig. 5A and B). The structure of the crystallographically determined virus could readily be superimposed onto the cryoEM map by aligning the icosahedral symmetry axes (Table 5). The cryoEM map had to be expanded by 5%, amounting to a radius increase of 6 Å, to obtain the best fit. This change in size of the EM map is well inside the error of determining the magnification of the electron microscope. The quality of the cryoEM map was excellent, as indicated by the resolution of the main chain in some places (Fig. 5E to G).

A 5.5-Å resolution difference map between the crystal structure electron density and the cryoEM map was calculated using the EMfit program (55). This showed two regions of density higher than three standard deviations of the background density, associated with the inside surface of the protein shell. The first region had an average height of about 6.5 σ and could be readily interpreted in terms of three nucleotides (Fig. 8A). This structure was an extended trinucleotide with the bases facing the inside capsid surface (Fig. 8B), close to the icosahedral 3-fold axes with the interaction with Tyr337 and Gln252. The second region had an average height of 3.5 σ but was not easily interpreted in terms of a standard nucleotide structure. Icosahedrally ordered genome structure has been previously observed in canine parvovirus (36) and minute virus of mice (35) but not in invertebrate densoviruses (Table 1).

The density in the channel along the 5-fold axes in the cryoEM map of the emptied particles was similar to that in the crystallographically determined map of the full infectious particle calculated to 5.5-Å resolution (Fig. 5C and D). Thus, the emptied particles still have the glycine-rich region occupying the channel along the 5-fold axes in at least some of the 12 channels of each particle. The externalization of the VP1 termini does not seem to have caused much damage to the particles. In contrast to the case for AdDNV, there is little density along the 5-fold axes of the shrimp (*PstDNV*) and silkworm (*BmDNV*) densoviruses. However, these structures are of recombinant particles that contain only VP3 or VP4, respectively, and are therefore missing the glycine-rich region. These observations pose the intriguing question of whether the PLA2 domain has to be refolded to be threaded through the 5-fold pore or whether the pore opens with restoration of the initial capsid structure after extrusion.

ACKNOWLEDGMENTS

We thank Agustin Avila-Sakar and Valorie Bowman for help with the cryo-electron microscopy. We are grateful to Sheryl Kelly for help in preparing the manuscript. We also thank the staff of beamline 23ID, GM/CA-CAT, at Advanced Photon Source, Argonne National Laboratory, for help with the data collection.

Use of the Advanced Photon Source was supported by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences, under contract DE-AC02-06CH11357. The work was supported by an NIH grant award (AI11219) to M.G.R. The work was also supported by Purdue University funds for Structural Biology and the Electron Microscope Facility. P.T. was supported by a grant from the Natural Sciences and Engineering Research Council of Canada, and Q.Y. acknowledges tuition waivers at INRS-Institut Armand-Frappier and a scholarship from the People's Republic of China.

G.M., X.Z., P.P., Q.Y., P.T., and M.G.R. designed research; G.M., X.Z., P.P., and Q.Y. performed research; G.M. analyzed data; Q.Y. and P.T. contributed reagents/analytic tools; and G.M., P.T., and M.G.R. wrote the paper.

We declare no conflict of interest.

REFERENCES

- Berns K, Parrish CR. 2006. *Parvoviridae*, p 2437–2478. In Knipe DM, Howley PM (ed), *Fields virology*. Lippincott Williams & Wilkins, Philadelphia, PA.
- Bergoin M, Tijssen P. 2000. Molecular biology of *Densovirinae*. *Contrib. Microbiol.* 4:12–32.
- Harvell CD, Kim K, Burkholder JM, Colwell RR, Epstein PR, Grimes DJ, Hofmann EE, Lipp EK, Osterhaus AD, Overstreet RM, Porter JW, Smith GW, Vasta GR. 1999. Emerging marine diseases—climate links and anthropogenic factors. *Science* 285:1505–1510.
- Li Y, Zadori Z, Bando H, Dubur R, Fediere G, Szelei J, Tijssen P. 2001. Genome organization of the densovirus from *Bombyx mori* (*BmDNV*) and enzyme activity of its capsid. *J. Gen. Virol.* 82:2821–2825.
- Liu K, Li Y, Jousset FX, Zadori Z, Szelei J, Yu Q, Pham HT, Lepine F, Bergoin M, Tijssen P. 2011. The *Acheta domesticus* densovirus, isolated from the European house cricket, has evolved an expression strategy unique among parvoviruses. *J. Virol.* 85:10069–10078.
- Szelei J, Woodring J, Goettel MS, Duke G, Jousset FX, Liu KY, Zadori Z, Li Y, Styler E, Boucias DG, Kleespies RG, Bergoin M, Tijssen P. 2011. Susceptibility of North-American and European crickets to *Acheta domesticus* densovirus (AdDNV) and associated epizootics. *J. Invertebr. Pathol.* 106:394–399.
- Carlson J, Suchman E, Buchatsky L. 2006. Densoviruses for control and genetic manipulation of mosquitoes. *Adv. Virus Res.* 68:361–392.
- Tijssen P. 1999. Molecular and structural basis of the evolution of parvovirus tropism. *Acta Vet. Hung.* 47:379–394.
- Canaan S, Zadori Z, Ghomashchi F, Bollinger J, Sadilek M, Moreau ME, Tijssen P, Gelb MH. 2004. Interfacial enzymology of parvovirus phospholipases A2. *J. Biol. Chem.* 279:14502–14508.

10. Girod A, Wobus CE, Zadori Z, Ried M, Leike K, Tijssen P, Klein-schmidt JA, Hallek M. 2002. The VP1 capsid protein of adeno-associated virus type 2 is carrying a phospholipase A2 domain required for virus infectivity. *J. Gen. Virol.* 83:973–978.
11. Zadori Z, Szelei J, Lacoste MC, Li Y, Gariepy S, Raymond P, Allaire M, Nabi IR, Tijssen P. 2001. A viral phospholipase A2 is required for parvovirus infectivity. *Dev. Cell* 1:291–302.
12. Tijssen P, Zadori Z. April 2002. Viral phospholipase A2 enzymes, anti-viral agents and methods of use. WO patent 2,002,000,924.
13. Farr GA, Zhang LG, Tattersall P. 2005. Parvoviral virions deploy a capsid-tethered lipolytic enzyme to breach the endosomal membrane during cell entry. *Proc. Natl. Acad. Sci. U. S. A.* 102:17148–17153.
14. Ros C, Gerber M, Kempf C. 2006. Conformational changes in the VP1-unique region of native human parvovirus B19 lead to exposure of internal sequences that play a role in virus neutralization and infectivity. *J. Virol.* 80:12017–12024.
15. Mani B, Baltzer C, Valle N, Almendral JM, Kempf C, Ros C. 2006. Low pH-dependent endosomal processing of the incoming parvovirus minute virus of mice virion leads to externalization of the VP1 N-terminal sequence (N-VP1), N-VP2 cleavage, and uncoating of the full-length genome. *J. Virol.* 80:1015–1024.
16. Venkatakrishnan B, Yarbrough J, Domsic J, Bennett A, Bothner B, Kozyreva OG, Samulski RJ, Muzychka N, McKenna R, Agbandje-McKenna M. 2013. Structure and dynamics of adeno-associated virus serotype 1 VP1-unique N-terminal domain and its role in capsid trafficking. *J. Virol.* 87:4974–4984.
17. Xie Q, Chapman MS. 1996. Canine parvovirus capsid structure, analyzed at 2.9 Å resolution. *J. Mol. Biol.* 264:497–520.
18. Agbandje M, McKenna R, Rossmann MG, Strassheim ML, Parrish CR. 1993. Structure determination of feline panleukopenia virus empty particles. *Proteins* 16:155–171.
19. Simpson AA, Hebert B, Sullivan GM, Parrish CR, Zadori Z, Tijssen P, Rossmann MG. 2002. The structure of porcine parvovirus: comparison with related viruses. *J. Mol. Biol.* 315:1189–1198.
20. Llamas-Saiz AL, Agbandje-McKenna M, Wikoff WR, Bratton J, Tattersall P, Rossmann MG. 1997. Structure determination of minute virus of mice. *Acta Crystallogr. D Biol. Crystallogr.* 53:93–102.
21. Halder S, Nam HJ, Govindasamy L, Vogel M, Dinsart C, Salome N, McKenna R, Agbandje-McKenna M. 2013. Structural characterization of H-1 parvovirus: comparison of infectious virions to empty capsids. *J. Virol.* 87:5128–5140.
22. Kaufmann B, Simpson AA, Rossmann MG. 2004. The structure of human parvovirus B19. *Proc. Natl. Acad. Sci. U. S. A.* 101:11628–11633.
23. Simpson AA, Chipman PR, Baker TS, Tijssen P, Rossmann MG. 1998. The structure of an insect parvovirus (*Galleria mellonella* densovirus) at 3.7 Å resolution. *Structure* 6:1355–1367.
24. Kaufmann B, El-Far M, Plevka P, Bowman VD, Li Y, Tijssen P, Rossmann MG. 2011. Structure of *Bombyx mori* densovirus I, a silkworm pathogen. *J. Virol.* 85:4691–4697.
25. Kaufmann B, Bowman VD, Li Y, Szelei J, Waddell PJ, Tijssen P, Rossmann MG. 2010. The structure of *Penaeus stylostris* densovirus, a shrimp pathogen. *J. Virol.* 84:11289–11296.
26. Xie Q, Bu W, Bhatia S, Hare J, Somasundaram T, Azzi A, Chapman MS. 2002. The atomic structure of adeno-associated virus (AAV-2), a vector for human gene therapy. *Proc. Natl. Acad. Sci. U. S. A.* 99:10405–10410.
27. Chapman MS, Agbandje-McKenna M. 2006. Atomic structure of viral particles, p 107–123. In Kerr JR, Cotmore S, Bloom ME, Linden RM, Parrish CR (ed), *Parvoviruses*. Hodder Arnold, London, United Kingdom.
28. Rossmann MG, Arnold E, Erickson JW, Frankenberger EA, Griffith JP, Hecht HJ, Johnson JE, Kamer G, Luo M, Mosser AG, et al. 1985. Structure of a human common cold virus and functional relationship to other picornaviruses. *Nature* 317:145–153.
29. Harrison SC, Olson AJ, Schutt CE, Winkler FK, Bricogne G. 1978. Tomato bushy stunt virus at 2.9 Å resolution. *Nature* 276:368–373.
30. Roberts MM, White JL, Grutter MG, Burnett RM. 1986. Three-dimensional structure of the adenovirus major coat protein hexon. *Science* 232:1148–1151.
31. Benson SD, Bamford JKH, Bamford DH, Burnett RM. 1999. Viral evolution revealed by bacteriophage PRD1 and human adenovirus coat protein structures. *Cell* 98:825–833.
32. Nandhagopal N, Simpson AA, Gurnon JR, Yan X, Baker TS, Graves MV, Van Etten JL, Rossmann MG. 2002. The structure and evolution of the major capsid protein of a large, lipid-containing DNA virus. *Proc. Natl. Acad. Sci. U. S. A.* 99:14758–14763.
33. Bahar MW, Graham SC, Stuart DI, Grimes JM. 2011. Insights into the evolution of a complex virus from the crystal structure of vaccinia virus D13. *Structure* 19:1011–1020.
34. Xiao C, Kuznetsov YG, Sun S, Hafenstein SL, Kostyuchenko VA, Chipman PR, Suzan-Monti M, Raoult D, McPherson A, Rossmann MG. 2009. Structural studies of the giant mimivirus. *PLoS Biol.* 7:e92. doi:10.1371/journal.pbio.1000092.
35. Agbandje-McKenna M, Llamas-Saiz AL, Wang F, Tattersall P, Rossmann MG. 1998. Functional implications of the structure of the murine parvovirus, minute virus of mice. *Structure* 6:1369–1381.
36. Tsao J, Chapman MS, Agbandje M, Keller W, Smith K, Wu H, Luo M, Smith TJ, Rossmann MG, Compans RW, Parrish CR. 1991. The three-dimensional structure of canine parvovirus and its functional implications. *Science* 251:1456–1464.
37. Chapman MS, Rossmann MG. 1993. Structure, sequence, and function correlations among parvoviruses. *Virology* 194:491–508.
38. Hogli JM, Chow M, Filman DJ. 1985. Three-dimensional structure of poliovirus at 2.9 Å resolution. *Science* 229:1358–1365.
39. Otwowski Z, Minor W. 1997. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* 276:307–326.
40. Tong L, Rossmann MG. 1997. Rotation function calculations with GLRF program. *Methods Enzymol.* 276:594–611.
41. Rossmann MG. 1990. The molecular replacement method. *Acta Crystallogr. A* 46:73–82.
42. Kleywegt GJ, Read RJ. 1997. Not your average density. *Structure* 5:1557–1569.
43. Collaborative Computational Project Number 4. 1994. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D Biol. Crystallogr.* 50:760–763.
44. Emsley P, Cowtan K. 2004. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* 60:2126–2132.
45. Brünger AT, Adams PD, Clore GM, DeLano WL, Gros P, Grosse-Kunstleve RW, Jiang JS, Kuszewski J, Nilges M, Pannu NS, Read RJ, Rice LM, Simonson T, Warren GL. 1998. Crystallography and NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr. D Biol. Crystallogr.* 54:905–921.
46. Ludtke SJ, Baldwin PR, Chiu W. 1999. EMAN: semiautomated software for high-resolution single-particle reconstructions. *J. Struct. Biol.* 128:82–97.
47. Baker ML, Zhang J, Ludtke SJ, Chiu W. 2010. Cryo-EM of macromolecular assemblies at near-atomic resolution. *Nat. Protoc.* 5:1697–1708.
48. Scheres SH, Chen S. 2012. Prevention of overfitting in cryo-EM structure determination. *Nat. Methods* 9:853–854.
49. Higgins DG, Sharp PM. 1988. CLUSTAL: a package for performing multiple sequence alignment on a microcomputer. *Gene* 73:237–244.
50. Rao ST, Rossmann MG. 1973. Comparison of super-secondary structures in proteins. *J. Mol. Biol.* 76:241–256.
51. Tamura K, Dudley J, Nei M, Kumar S. 2007. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol. Biol. Evol.* 24:1596–1599.
52. Garriga D, Pickl-Herk A, Luque D, Wruss J, Caston JR, Blaas D, Verdaguera N. 2012. Insights into minor group rhinovirus uncoating: the X-ray structure of the HRV2 empty capsid. *PLoS Pathog.* 8:e1002473. doi:10.1371/journal.ppat.1002473.
53. Boisvert M, Fernandes S, Tijssen P. 2010. Multiple pathways involved in porcine parvovirus cellular entry and trafficking toward the nucleus. *J. Virol.* 84:7782–7792.
54. Huotari J, Helenius A. 2011. Endosome maturation. *EMBO J.* 30:3481–3500.
55. Rossmann MG, Bernal R, Pletnev SV. 2001. Combining electron microscopic with X-ray crystallographic structures. *J. Struct. Biol.* 136:190–200.
56. Wu H, Rossmann MG. 1993. The canine parvovirus empty capsid structure. *J. Mol. Biol.* 233:231–244.

Organization of the Ambisense Genome of the *Helicoverpa armigera* Densovirus

Mohamed El-Far,* Jozsef Szelei, Qian Yu, Gilles Fédière,* Max Bergoin, and Peter Tijssen

INRS-Institut Armand-Frappier, Laval, Quebec, Canada

A natural densovirus (DNV) of a serious phytophagous pest, *Helicoverpa armigera*, was isolated. The genome of HaDNV contained 6,039 nucleotides (nt) and included inverted terminal repeats (ITRs) of 545 nt with terminal Y-shaped hairpins of 126 nt. Its DNA sequence and ambisense organization with four typical open reading frames (ORFs) demonstrated that it belonged to the genus *Densovirus* in the subfamily *Densovirinae* of the family *Parvoviridae*.

The cotton bollworm, *Helicoverpa armigera* (Hübner) (Noctuidae), is one of the most serious polyphagous pest species, causing huge losses predominantly in the Old World. Economic damage is greatest in cotton and vegetables. In grain legumes, which are staple foods for people in many countries, up to 80% of the crop can be destroyed. The cotton bollworm also attacks a great number of cereal, vegetable, and garden crops.

The cotton bollworm has become increasingly resistant to pesticides (3), necessitating biological or integrated pest management. Densoviruses (DNVs) have shown great potential, e.g., complete control of *Casphalia extranea* in the oil palm industry in Ivory Coast (1). Since the late 1980s, we have been developing a program to obtain natural DNV isolates from pests in agricultural fields in Egypt. This program led previously to the isolation of the MIDNV (a pathogen of the maize worm) (2). Here we report the isolation, cloning, and sequence analysis of a densovirus, HaDNV, from the cotton bollworm. Infected larvae stopped feeding within a few days and died within a week. The virus remained infectious for months in cadavers in the field.

HaDNV DNA, extracted from the virus under conditions of high ionic strength to anneal the single-stranded DNA (ssDNA), has a size of around 6 kb. This DNA was blunt-ended by a mixture of Klenow fragment and T4 DNA polymerase and cloned into the EcoRV site of the pBluescript vector. The restriction profiles of the viral DNA indicated a close relationship to both MIDNV and *Galleria mellonella* densovirus (GmDNV). Digesting viral DNA with BamHI gave rise to a 5.5-kb fragment and a 0.3-kb fragment. This small DNA fragment was probably a doublet, resulting from the existence of symmetrical BamHI sites within the viral putative inverted terminal repeats (ITRs), similar to fragments from some other densoviruses (2, 4, 5). This observation helped to clone the viral DNA by methods that were successfully employed for GmDNV and MIDNV (2, 5). Two complete clones, pHa2 and pHa8, were sequenced, both in both directions, using Sanger's method and the primer-walking method as described before (5). The sequences of the two clones were identical.

The HaDNV genome contained inverted terminal repeats (ITRs), typical of members of the *Densovirus* genus, with a length of 545 nucleotides (nt). The terminal Y-shaped hairpins of 126 nt are completely conserved between HaDNV and MIDNV. The overall sequence is 90% identical to the corresponding sequence of MIDNV and 86% identical to that of GmDNV.

The large open reading frame (ORF) 1 (nt 657 to 1352) on the plus strand had a coding capacity for NS1 of 547 amino acids (aa). ORF2 corresponded to NS2 (nt 1359 to 2996) with 275 aa, and

ORF3 (nt 1366 to 2193) corresponded to NS3 with 231 aa. On the complementary minus strand, a large ORF (also on the 5' half at nt 3028 to 5463) with a potential coding region of 811 aa corresponded well to those of the VP structural proteins of related densoviruses. The distribution of the putative coding sequences of HaDNV on the two viral strands thus implied an ambisense organization, and HaDNV contained the typical NS-1 helicase superfamily III and VP phospholipase A2 (6) motifs observed in parvoviruses.

Nucleotide sequence accession number. The GenBank accession number of HaDNV is JQ894784.

ACKNOWLEDGMENTS

This work was supported by the Natural Sciences and Engineering Research Council of Canada to P.T. M.E.-F. acknowledges financial support and tuition waivers at INRS. Q.Y. acknowledges support from a scholarship from the People's Republic of China. G.F. was supported by IRD during his sabbatical at INRS. G.F. and M.B. are invited professors at INRS.

REFERENCES

1. Fediere G. 1996. Recherches sur des viroses lepidopteres ravageurs des cultures perennes en Côte d'Ivoire et de cultures annuelles en Egypte. Université Montpellier II, Montpellier, France.
2. Fediere G, El-Far M, Li Y, Bergoin M, Tijssen P. 2004. Expression strategy of densovirus from *Mythimna loreyi*. Virology 320:181–189.
3. Gunning RV, Dang HT, Kemp FC, Nicholson IC, Moores GD. 2005. New resistance mechanism in *Helicoverpa armigera* threatens transgenic crops expressing *Bacillus thuringiensis* Cry1Ac toxin. Appl. Environ. Microbiol. 71:2558–2563.
4. Tijssen P, et al. 2011. Parvoviridae, p 375–395. In King AMQ, Adams MJ, Carstens E, Lefkowitz EJ (ed), Virus taxonomy: classification and nomenclature of viruses; ninth report of the International Committee on Taxonomy of Viruses. Elsevier, San Diego, CA.
5. Tijssen P, et al. 2003. Organization and expression strategy of the ambisense genome of densovirus of *Galleria mellonella*. J. Virol. 77:10357–10365.
6. Zadori Z, et al. 2001. A viral phospholipase A2 is required for parvovirus infectivity. Dev. Cell 1:291–302.

Received 5 April 2012 Accepted 9 April 2012

Address correspondence to Peter Tijssen, peter.tijssen@iaf.inrs.ca.

* Present address: Mohamed El-Far, Boehringer Ingelheim Canada, Laval, Quebec, Canada; Gilles Fédière, IRD, Nouméa, New Caledonia.

Copyright © 2012, American Society for Microbiology. All Rights Reserved.

doi:10.1128/JVI.00865-12

Pseudoplusia includens Densovirus Genome Organization and Expression Strategy

Oanh T. H. Huynh, Hanh T. Pham, Qian Yu, and Peter Tijssen

INRS-Institut Armand-Frappier, Laval, Quebec, Canada

The genome of a densovirus of a major phytophagous pest, *Pseudoplusia includens*, was analyzed. It contained 5,990 nucleotides (nt) and included inverted terminal repeats of 540 nt with terminal Y-shaped hairpins of 120 nt. Its DNA sequence and ambisense organization with 4 typical open reading frames demonstrated that it belonged to the genus *Densovirus* in the subfamily *Densovirinae* of the family *Parvoviridae*.

The distribution of the polyphagous soybean looper pest, *Pseudoplusia includens* (syn., *Chrysodeixis includens* [Hübner] [Noctuidae, Plusiinae, Lepidoptera]), is restricted to the Western Hemisphere, occurring from southern Canada to southern South America (1). In addition to the soybean, it may feed on a large number of crops of economic importance (8, 9). Previously, two small icosahedral viruses have been isolated from the soybean looper, a picornavirus and a smaller virus with biophysical properties that seem to match those of the densoviruses (2).

Densovirus are notoriously unstable upon cloning (7, 10–13), and densovirus entries in GenBank, such as those from *Junonia coenia* (JcDNV) (3) and *Diatraea saccharalis* (DsDNV) (NC_001899), often lack significant parts of their inverted terminal repeats (ITRs). DNA purified from *Pseudoplusia includens* DNV (PiDNV) in phosphate-buffered saline (PBS) had a size of around 6 kb. This DNA was blunt ended by a mixture of Klenow fragment and T4 DNA polymerase and cloned into a linear pfazz vector (from Lucigen Corp.), which lacks transcription into the insert and torsional stress (5) to prevent recombination and deletion of insert fragments. Six clones, or about 0.3%, had full-length inserts and could be stably subcloned into circular vectors.

Four complete clones were sequenced in both directions, using Sanger's method and the primer-walking method as described before (11), and the contigs were assembled by the CAP3 program (<http://pbil.univ-lyon1.fr/cap3.php/>) (6). The difficulties encountered with sequencing of the terminal hairpins were solved by sequencing after (i) digestion near the middle of the hairpin with BstUI restriction enzyme or (ii) amplifying the hairpins by PCR in the presence three additives: 1.3 M betaine, 5% dimethyl sulfoxide, and 50 mM 7-deaza-dGTP. Sequences of the clones, except for the flip-flop regions in the hairpins, were identical. In the hairpins, nucleotides (nt) 46 to 75 and nt 5916 to 5945 occurred in two orientations, "flip" and its reverse complement orientation "flop." The ambisense PiDNV genome contained typical ITRs of members of the *Densovirus* genus with a length of 540 nt and terminal Y-shaped hairpins of 120 nt. The overall sequence of 5,990 nt was 83 to 87% identical with those of other viruses in the *Densovirus* genus but about 50 nt shorter.

The open reading frames (ORFs) were conserved with members of the *Densovirus* genus, and the putative splicing sites were conserved with those that have been identified for *Galleria mellonella* DNV (GmDNV) (11) and *Mythimna loreyi* DNV (MIDNV) (4). The large ORF1 (nt 1355 to 3019) on the plus strand had a coding capacity for NS1 of 554 amino acids (aa), ORF2 corre-

sponded to NS2 (nt 1362 to 3019) with 275 aa, and ORF3 (nt 647 to 1348) corresponded to NS3 with 233 aa. On the complementary minus strand, a large ORF (also on the 5' half at nt 3006 to 5423) with a potential coding region of 805 aa corresponded well to those of the VP structural proteins of related densoviruses. The distribution of the putative coding sequences implied an ambisense organization and expression, and PiDNV contained the typical NS-1 helicase superfamily III and VP phospholipase A2 (14) motifs observed in other parvoviruses.

Nucleotide sequence accession number. The GenBank accession number of PiDNV is [JX645046](#).

ACKNOWLEDGMENTS

This work was supported by the Natural Sciences and Engineering Research Council of Canada to P.T.O.T.H. acknowledges support from a scholarship from the Biochemical Institute of Ho Chi Minh City, Vietnam. Q.Y. acknowledges support from a scholarship from the People's Republic of China. O.T.H.H., H.T.P., and Q.Y. acknowledge tuition waivers at INRS.

REFERENCES

1. Alford RA, Hammond AM, Jr. 1982. Plusiinae (Lepidoptera: Noctuidae) populations in Louisiana soybean ecosystems as determined with phloore-baited traps. *J. Econ. Entomol.* 75:647–650.
2. Chao Y-C, Young SY III, Kim KS, Scott HA. 1985. A newly isolated densoencelosis virus from *Pseudoplusia includens* (Lepidoptera: Noctuidae). *J. Invertebr. Pathol.* 46:70–82.
3. Dumas B, Jourdan M, Pascaud AM, Bergoin M. 1992. Complete nucleotide sequence of the cloned infectious genome of *Junonia coenia* densoivirus reveals an organization unique among parvoviruses. *Virology* 191: 202–222.
4. Fediere G, El-Far M, Li Y, Bergoin M, Tijssen P. 2004. Expression strategy of densoencelosis virus from *Mythimna loreyi*. *Virology* 320:181–189.
5. Godiska R, et al. 2010. Linear plasmid vector for cloning of repetitive or unstable sequences in *Escherichia coli*. *Nucleic Acids Res.* 38:e88.
6. Huang X, Madan A. 1999. CAP3: a DNA sequence assembly program. *Genome Res.* 9:868–877.
7. Liu K, et al. 2011. The *Acheta domesticus* densovirus, isolated from the European house cricket, has evolved an expression strategy unique among parvoviruses. *J. Virol.* 85:10069–10078.

Received 10 September 2012 Accepted 11 September 2012

Address correspondence to Peter Tijssen, peter.tijssen@af.inrs.ca.

Copyright © 2012, American Society for Microbiology. All Rights Reserved.

[doi:10.1128/JVI.02462-12](https://doi.org/10.1128/JVI.02462-12)

8. MacRae TC, et al. 2005. Laboratory and field evaluations of transgenic soybean exhibiting high-dose expression of a synthetic *Bacillus thuringiensis* cry1A gene for control of Lepidoptera. *J. Econ. Entomol.* 98:577–587.
9. McPherson RM, MacRae TC. 2009. Evaluation of transgenic soybean exhibiting high expression of a synthetic *Bacillus thuringiensis* cry1A transgene for suppressing lepidopteran population densities and crop injury. *J. Econ. Entomol.* 102:1640–1648.
10. Tijssen P, et al. 2006. Evolution of densoviruses, p 55–68. In Kerr JR, Cotmore SF, Bloom ME, Linden RM, Parrish CR (ed), *Parvoviruses*. Hodder Arnold, London, United Kingdom.
11. Tijssen P, et al. 2003. Organization and expression strategy of the ambisense genome of densonucleosis virus of *Galleria mellonella*. *J. Virol.* 77:10357–10365.
12. Yu Q, Fediere G, Abd-Alla A, Bergoin M, Tijssen P. 2012. Iteravirus-like genome organization of a densovirus from *Sibine fusca* Stoll. *J. Virol.* 86:8897–8898.
13. Yu Q, Hajek AE, Bergoin M, Tijssen P. 2012. *Papilio polyxenes* densovirus has an iteravirus-like genome organization. *J. Virol.* 86:9534–9535.
14. Zadori Z, et al. 2001. A viral phospholipase A2 is required for parvovirus infectivity. *Dev. Cell* 1:291–302.

A Novel Ambisense Densovirus, *Acheta domesticus* Mini Ambidensovirus, from Crickets

Hanh T. Pham, Qian Yu, Max Bergoin, Peter Tijssen

INRS-Institut Armand-Frappier, Laval, QC, Canada

The genome structure of *Acheta domesticus* mini ambidensovirus, isolated from crickets, resembled that of ambisense densoviruses from *Lepidoptera* but was 20% smaller. It had the highest (<25%) protein sequence identity with the nonstructural protein 1 (NS1) of *Iteravirus* and VP of *Densovirus* members (both with 25% coverage) and smaller (0.2- versus 0.55-kb) Y-shaped inverted terminal repeats.

Received 30 September 2013 Accepted 2 October 2013 Published 7 November 2013

Citation Pham HT, Yu Q, Bergoin M, Tijssen P. 2013. A novel ambisense densovirus, *Acheta domesticus* mini ambidensovirus, from crickets. *Genome Announc.* 1(6):e00914-13. doi:10.1128/genomeA.00914-13.

Copyright © 2013 Pham et al. This is an open-access article distributed under the terms of the Creative Commons Attribution 3.0 Unported license.

Address correspondence to Peter Tijssen, peter.tijssen@iaf.inrs.ca.

The cricket industry has been devastated worldwide recently by the *Acheta domesticus* densovirus (AdDNV) (1–4). We also observed several, thus far unknown, viruses such as parvoviruses, which have circular, single-stranded DNA (ssDNA) genomes (5), and a new densovirus (parvovirus).

Two genera of insect parvoviruses, named densoviruses (6), are particularly relevant for this new densovirus. The *Densovirus* genus contains ambisense densoviruses from *Lepidoptera*, with genomes of 6 kb, Y-shaped inverted terminal repeats (ITRs) of about 0.55 kb, and sequence identities of about 85% (7–11). The *Iteravirus* genus contains monosense densoviruses, also from *Lepidoptera*, with 5-kb genomes, J-shaped 0.25-kb inverted terminal repeats (ITRs), and about 75% sequence identities (12–15).

A new virus with morphology and size similar to densoviruses was detected in some cricket samples from the United States. Virus was purified and DNA extracted as described previously (5). Digestion of viral DNA with EcoRI yielded 2 bands of about 700 bp and 4,200 bp on agarose gels. DNA was blunt ended with T4 DNA polymerase and a large Klenow fragment in the presence of dNTPs at room temperature (RT), ligated into the EcoRV site of pBluescript KS(+), and transformed into SURE cells. DNA of clones with expected sizes were subcloned. Digestion with EcoNI within the terminal hairpins yielded clear reads of ITR sequences. Several complete clones were sequenced in both directions by use of Sanger's primer-walking method as described previously (11). Contigs were assembled by use of the CAP3 program (<http://pbil.univ-lyon1.fr/cap3.php/>) (16).

Surprisingly, the genome structure and gene organization of this virus strongly resembled those of ambisense densoviruses from the *Densovirus* genus (7–11), but the genome sequence was only 4,945 nucleotides (nt) long, instead of about 6,035 nt, and lacked nucleotide sequence identity (best E value of 0.017, with a query coverage of 1%). Protein sequence identities were for the major nonstructural protein 1 (NS1) closest to *Iteravirus* members and, oddly, for the structural proteins (VP) closest to *Densovirus* members (both at best 25% identity for 25% coverage [or higher for shorter coverage]).

ITRs of AdMADV were smaller than those of densovirus members (199 versus about 545 nt) and Y-shaped, with a 113-nt hairpin. The 45-nt-long stem contained two side arms in the middle, nt 46 to 68, that occurred in two sequence orientations (flip/flop). It had a high GC content (63%) and contained inboard TATA boxes, at 193 to 199 for the NS cassette and at 4747 to 4753 for the VP cassette. This structure is identical to that of *Densovirus* ITRs.

The NS cassette consisted, as for *Densovirus* members, of NS3, followed by NS1 and an overlapping NS2. Splicing, as for *Densovirus*, would remove the NS3 open reading frame (ORF) and allow expression of NS1 and NS2 by leaky scanning. As for *Densovirus*, the putative splice acceptor site was located just upstream of the initiation codon of NS1 (1172-CAG/aATG_{NS1..N_{19..ATG_{NS2}}}) (in GmDNV, 1395-CAG/ATG_{NS1..N_{4..ATG_{NS2}}}). As for members of the *Densovirus* genus, the VP on the complementary strand also contained the phospholipase A2 motif (4,590 to 4,680 nt) (17) and the stop codons of NS1 and VP were neighbors (2661-TAG/AAT-2666), suggesting a small overlap of their transcripts, as for GmDNV.

Nucleotide sequence accession number. The GenBank accession number of AdMADV is [KF275669](https://www.ncbi.nlm.nih.gov/nuccore/KF275669).

ACKNOWLEDGMENTS

This work was supported by the Natural Sciences and Engineering Research Council of Canada to P.T.; H.T.P. and Q.Y. acknowledge tuition fee waivers at INRS and scholarships from INRS and the People's Republic of China.

REFERENCES

1. Liu K, Li Y, Jousset FX, Zadori Z, Szelei J, Yu Q, Pham HT, Lépine F, Bergoin M, Tijssen P. 2011. The *Acheta domesticus* densovirus, isolated from the European house cricket, has evolved an expression strategy unique among parvoviruses. *J. Virol.* 85:10069–10078.
2. Szelei J, Woodring J, Goettel MS, Duke G, Jousset FX, Liu KY, Zadori Z, Li Y, Styer E, Boucias DG, Kleespies RG, Bergoin M, Tijssen P. 2011. Susceptibility of North-American and European crickets to *Acheta domesticus* densovirus (AdDNV) and associated epizootics. *J. Invertebr. Pathol.* 106:394–399.
3. Weissman DB, Gray DA, Pham HT, Tijssen P. 2012. Billions and billions

- sold: pet-feeder crickets (*Orthoptera: Gryllidae*), commercial cricket farms, an epizootic densovirus, and government regulations make a potential disaster. Zootaxa 3504:67–88.
4. Pham HT, Iwao H, Szelei J, Li Y, Liu K, Bergoin M, Tijssen P. 2013. Comparative genomic analysis of *Acheta domesticus* densovirus from different outbreaks in Europe, North America, and Japan. Genome Announc. 1(4):e00629-13. doi:[10.1128/genomeA.00629-13](https://doi.org/10.1128/genomeA.00629-13).
 5. Pham HT, Bergoin M, Tijssen P. 2013. *Acheta domesticus* Volvovirus, a novel single-stranded circular DNA virus of the house cricket. Genome Announc. 1(2):e00079-13. doi:[10.1128/genomeA.00079-13](https://doi.org/10.1128/genomeA.00079-13).
 6. Tijssen P, Agbandje-McKenna M, Almendral JM, Bergoin M, Flegel TW, Hedman K, Kleinschmidt JA, Li Y, Pintel DJ, Tattersall P. 2011. Parvoviridae, p 375–395. In King AMQ, Adams MJ, Carstens E, Lefkowitz EJ (ed), Virus taxonomy: classification and nomenclature of viruses: ninth report of the International Committee on Taxonomy of Viruses. Elsevier, San Diego, CA.
 7. Pham HT, Nguyen OT, Jousset FX, Bergoin M, Tijssen P. 2013. *Junonia coenia* densovirus (JcDNV) genome structure. Genome Announc. 1(4):e00591-13. doi:[10.1128/genomeA.00591-13](https://doi.org/10.1128/genomeA.00591-13).
 8. El-Far M, Szelei J, Yu Q, Fedière G, Bergoin M, Tijssen P. 2012. Organization of the ambisense genome of the *Helicoverpa armigera* densovirus. J. Virol. 86:7024. doi:[10.1128/JVI.00865-12](https://doi.org/10.1128/JVI.00865-12).
 9. Huynh OT, Pham HT, Yu Q, Tijssen P. 2012. *Pseudoplusia includens* densovirus genome organization and expression strategy. J. Virol. 86: 13127–13128.
 10. Fédière G, El-Far M, Li Y, Bergoin M, Tijssen P. 2004. Expression strategy of densonucleosis virus from *Mythimna loreyi*. Virology 320: 181–189.
 11. Tijssen P, Li Y, El-Far M, Szelei J, Letarte M, Zádori Z. 2003. Organization and expression strategy of the ambisense genome of densonucleosis virus of *Galleria mellonella*. J. Virol. 77:10357–10365.
 12. Yu Q, Hajek AE, Bergoin M, Tijssen P. 2012. *Papilio polyxenes* densovirus has an iteravirus-like genome organization. J. Virol. 86:9534–9535.
 13. Yu Q, Fédière G, Abd-Alla A, Bergoin M, Tijssen P. 2012. Iteravirus-like genome organization of a densovirus from *Sibine fusca* Stoll. J. Virol. 86: 8897–8898.
 14. Fédière G, Li Y, Zádori Z, Szelei J, Tijssen P. 2002. Genome organization of *Casphalia extranea* densovirus, a new iteravirus. Virology 292:299–308.
 15. Li Y, Zádori Z, Bando H, Dubuc R, Fédière G, Szelei J, Tijssen P. 2001. Genome organization of the densovirus from *Bombyx mori* (BmDNV-1) and enzyme activity of its capsid. J. Gen. Virol. 82:2821–2825.
 16. Huang X, Madan A. 1999. CAP3: A DNA sequence assembly program. Genome Res. 9:868–877.
 17. Zádori Z, Szelei J, Lacoste MC, Li Y, Gariépy S, Raymond P, Allaire M, Nabi IR, Tijssen P. 2001. A viral phospholipase A2 is required for parvovirus infectivity. Dev. Cell 1:291–302.

A Circo-Like Virus Isolated from *Penaeus monodon* Shrimps

Hanh T. Pham,^a Qian Yu,^a Maude Boisvert,^a Hanh T. Van,^b Max Bergoin,^a Peter Tijssen^a

INRS, Institut Armand-Frappier, Laval, QC, Canada^a; Institute of Tropical Biology, Vietnam Academy of Science and Technology, Ho Chi Minh City, Vietnam^b

A virus with a circular Rep-encoding single-stranded DNA (ssDNA) (CRESS-DNA) genome (PmCV-1) was isolated from *Penaeus monodon* shrimps in Vietnam. The gene structure of the 1,777-nucleotide (nt) genome was similar to that of circo-viruses and cycloviruses, but the nucleic acid and protein sequence identities to these viruses were very low.

Received 5 December 2013 Accepted 7 December 2013 Published 16 January 2014

Citation Pham HT, Yu Q, Boisvert M, Van HT, Bergoin M, Tijssen P. 2014. A circo-like virus isolated from *Penaeus monodon* shrimps. Genome Announc. 2(1):e01172-13. doi: 10.1128/genomeA.01172-13.

Copyright © 2014 Pham et al. This is an open-access article distributed under the terms of the Creative Commons Attribution 3.0 Unported license. Address correspondence to Peter Tijssen, peter.tijssen@iaf.inrs.ca.

Recently, viral metagenomics revealed circo-like viruses in the marine copepod species *Acartia tonsa* and *Labidocera aestiva* (Crustaceae) (1). Here, we report the isolation by classical methods of a similar virus with a circular Rep-encoding single-stranded DNA (ssDNA) (CRESS-DNA) genome from *Penaeus monodon* shrimps (PmCV-1).

Circoviruses are nonenveloped, icosahedral particles and contain circular ssDNA genomes of about 1.7 to 2.3 kb. The open reading frame (ORF) for the Rep protein codes for conserved rolling circle replication (RCR) and superfamily 3 (SF3) helicase motifs (2, 3). In contrast, the cap gene is generally not conserved. Originally, circoviruses were isolated from pig and bird species (4–6), but *in vitro* rolling circle amplification, high-throughput sequencing, and metagenomics studies have led to rapid expansion of the known diversity and host range of small CRESS-DNA viruses (CVs). This also led to an unsettled viral taxonomy with different subfamilies within the *Circoviridae* family and reassignment of their members (2).

In this study, about 100 g of cleaned, diseased *Penaeus monodon* shrimps was homogenized and virus was purified (isolate VN11 from Vietnam) as described previously (7, 8). Viral DNA was isolated from purified viruses with the High Pure viral nucleic acid kit (Roche Applied Science), followed by rolling circle amplification by φ29 DNA polymerase (NEB) at 30°C for 6 hours (9). Amplified product was then digested with EcoRI and separated on a 1% agarose gel. A band of 1.8 kb was recovered from the gel and cloned into a pBluescriptKS(+) vector. Clones were sequenced by Sanger's method and primer walking. PCR with outward primers was carried out and the amplicon was cloned into a TA vector (pGEMT-easy, Promega). All sequencing results were assembled using the CAP3 program (10).

Sequence analysis revealed that PmCV-1 is closely related to members of the *Cyclovirus* genus in the *Circoviridae* family. PmCV-1 possesses a 1,777-nucleotide (nt) genome containing three ORFs encoding 266, 255, and 146 amino acids (aa). Numbering starts with the loop in the conserved stem loop. The 266-aa product of the largest ORF, from nt 51 to 851, shared about 30% sequence identity (over 90% of query coverage) with the putative Rep of cycloviruses and contained RCR and SF3 motifs. The 255-aa product of the ORF translated in the opposite di-

rection, from nt 1,671 to 904, shared 25% identity with the Cap protein of a *Diporeia* sp.-associated circular virus (GenBank accession no. KC248415.1, E value 0.004), and thus the ORF probably encodes the capsid protein. The smallest ORF, from nt 1,246 to 1,686, codes for a 146-aa protein that did not reveal any amino acid similarity using Blastx in a protein database with E values of <0.01. The 156-nt intergenic region between the 5' ends of putative cap and rep genes encompasses 13-nt inverted repeats (nt 11 to 23 and 1765 to 1777) forming a stem and a 10-nt loop containing a canonical nonanucleotide, TAATATTAC, between nt 2 and 10. The intergenic region between the 3' ends of the cap and rep genes is 53 nt long. The genome structure resembles that of circoviruses and cycloviruses.

Metagenomic discovery has particularly impacted the discovery of CRESS-DNA viruses, both in host range and genetic diversity. Although this approach is very powerful, its perils should not be underestimated (11).

Nucleotide sequence accession number. The GenBank accession number for PmCV-1 is [KF481961](#).

ACKNOWLEDGMENTS

This work was supported by the Natural Sciences and Engineering Research Council of Canada (NSERC) to P.T.; H.T.P. and Q.Y. acknowledge tuition fee waivers at INRS and scholarships from INRS and the People's Republic of China. M.B. acknowledges a scholarship from NSERC.

REFERENCES

1. Dunlap DS, Ng TF, Rosario K, Barbosa JG, Greco AM, Breitbart M, Hewson I. 2013. Molecular and microscopic evidence of viruses in marine copepods. Proc. Natl. Acad. Sci. U. S. A. 110:1375–1380. <http://dx.doi.org/10.1073/pnas.1216595110>.
2. Rosario K, Duffy S, Breitbart M. 2012. A field guide to eukaryotic circular single-stranded DNA viruses: insights gained from metagenomics. Arch. Virol. 157:1851–1871. <http://dx.doi.org/10.1007/s00705-012-1391-y>.
3. Delwart E, Li L. 2012. Rapidly expanding genetic diversity and host range of the *Circoviridae* viral family and other Rep encoding small circular ssDNA genomes. Virus Res. 164:114–121. <http://dx.doi.org/10.1016/j.virusres.2011.11.021>.
4. Ritchie BW, Niagro FD, Lukert PD, Steffens WL, III, Latimer KS. 1989.

- Characterization of a new virus from cockatoos with psittacine beak and feather disease. *Virology* 171:83–88. [http://dx.doi.org/10.1016/0042-6822\(89\)90513-8](http://dx.doi.org/10.1016/0042-6822(89)90513-8).
5. Tischer I, Mields W, Wolff D, Vagt M, Griem W. 1986. Studies on epidemiology and pathogenicity of porcine circovirus. *Arch. Virol.* 91: 271–276. <http://dx.doi.org/10.1007/BF01314286>.
 6. Todd D, Niagro FD, Ritchie BW, Curran W, Allan GM, Lukert PD, Latimer KS, Steffens WL III, McNulty MS. 1991. Comparison of three animal viruses with circular single-stranded DNA genomes. *Arch. Virol.* 117:129–135. <http://dx.doi.org/10.1007/BF01310498>.
 7. Pham HT, Iwao H, Bergoin M, Tijssen P. 2013. New volvovirus isolates from *Acheta domesticus* (Japan) and *Gryllus assimilis* (United States). *Genome Announc.* 1(3):e00328-13. <http://dx.doi.org/10.1128/genomeA.00328-13>.
 8. Pham HT, Bergoin M, Tijssen P. 2013. *Acheta domesticus* volvovirus, a novel single-stranded circular DNA virus of the house cricket. *Genome Announc.* 1(2):e00079-13. <http://dx.doi.org/10.1128/genomeA.00079-13>.
 9. de Vega M, Lázaro JM, Mencía M, Blanca L, Salas M. 2010. Improvement of φ29 DNA polymerase amplification performance by fusion of DNA binding motifs. *Proc. Natl. Acad. Sci. U. S. A.* 107:16506–16511. <http://dx.doi.org/10.1073/pnas.1011428107>.
 10. Huang X, Madan A. 1999. CAP3: a DNA sequence assembly program. *Genome Res.* 9:868–877. <http://dx.doi.org/10.1101/gr.9.9.868>.
 11. Naccache SN, Greninger AL, Lee D, Coffey LL, Phan T, Rein-Weston A, Aronsohn A, Hackett J, Jr, Delwart EL, Chiu CY. 2013. The perils of pathogen discovery: origin of a novel parvovirus-like hybrid genome traced to nucleic acid extraction spin columns. *J. Virol.* 87:11966–11977. <http://dx.doi.org/10.1128/JVI.02323-13>.