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

Molecular biology of single-stranded DNA viruses in shrimps and crickets

Par

HANH THI PHAM

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

Jury d'évaluation

Président du jury et	**PATRICK LABONTÉ, Ph.D.**
examinateur interne	INRS-Institut Armand-Frappier
Examinateur externe	**CARL A. GAGNON, DMV, Ph.D.**
	Université de Montréal
Examinateur externe	**BENOÎT BARBEAU, Ph.D.**
	Université du Québec à Montréal
Directeur de recherche	**PETER TIJSSEN, Ph.D.**
	INRS-Institut Armand-Frappier
Co-directeur de recherche	**MAX BERGOIN, Ph.D.**
	Université Montpellier II, France

To my beloved family

whose endless love and unconditional sacrifice have inspired me to reach the end of my PhD journey.

Acknowledgements

My thesis would not have been completed without tremendous support and encouragement from many kind persons. Foremost, my supervisor, Professor Dr. **Peter Tijssen**, he has been guiding me into the real science world, giving me priceless support and independence to pursue various ideas and projects. He also offered me opportunities to participate in many international conferences during the years of my PhD. I did learn a lot from him, aside from his immense knowledge in science; I really admire his patience, generosity, and positive attitude. Going through major roadblocks, when I was slipping out of hope, I could feel his steadfast confidence in my abilities. I would like to express my deepest gratitude to him for believing in me and for all of his support.

I greatly appreciate my co-supervisor, Professor Dr. **Max Bergoin**, from the "Laboratoire de Pathologie Comparée, Université Montpellier II, France". He has supplied lots of materials as well as many great ideas for our research. He has been always available to correct all the manuscripts and my thesis.

I would like to acknowledge all financial support from the Agence universitaire de la Francophonie (AUF), Fondation Armand-Frappier, Discovery grant NSERC (Peter Tijssen), Programme d'exemption des frais de scolarité supplémentaires and from INRS-Institut Armand-Frappier.

I would also like to thank the Professors Dr. **Patrick Labonté** (INRS-Institut Armand-Frappier), Dr. **Carl A. Gagnon** (Université de Montréal), and Dr. **Benoît Barbeau** (Université du Québec à Montréal) who have been kindly sharing their time and helpful ideas for the accomplishment of this thesis.

I would like to express my sincere appreciation to Professor Dr. **Jonathan Perreault** for his enthusiastic support during my experiments with radioactive P^{32}/RPA. He has been very generous of his precious time to explain and answer all my questions about RNA techniques and to share his P^{32}/RPA experience. His support has been very valuable to the success of our research.

I am also grateful to Dr. **Jozsef Slezei** who provided me useful suggestions about the research, especially during my first years in the lab. I thank **Micheline Letarte** for her EM technical support. Many thanks to all my present and past lab colleagues for their various support, particularly **Sandra Fernandes** for her generosity in offering her help, **Maude**

Boisvert and **Qian Yu** for their lab responsibilities and kind collaboration. For me, it is my privilege to have been working with them.

I also want to acknowledge Dr. **Hanh Thi Van** (VAST-Institute of Tropical Biology in Hochiminh City, Vietnam) for her help in collecting infected shrimp samples from Vietnam. My sincere thankfulness is reserved for **Madame Anne Philippon,** for her administrative help.

It is my honour to know **Dr. Thérèse Trinh Huynh** (CHU Sainte-Justine) and **Dr. Quynh Nguyen** (VAST-Institute of Tropical Biology, Vietnam). I am thankful to have them as my great friends and my motivating, caring mentors who were always willing to share their experience and advice without hesitation.

Special thanks to my best friend, **Thuy Thanh Nguyen** who is always available to listen, share her opinions and encourage me. Her genuine friendship is irreplaceable.

Finally, to my dear husband, **Etienne Marchildon** and his family, thank you very much for supporting and encouraging me throughout this experience.

Table of contents

Preface

My Ph.D. project includes different aspects of research on single-stranded DNA viruses that infect shrimp and cricket. My initial goals focused on the molecular biology of *Penaeus stylirostris* densovirus (PstDNV), a shrimp densovirus. It is one of the most problematic viruses impacting worldwide aquaculture of shrimps due to the severe economic consequences that have been reported over several decades.

A major problem I faced was the minute amount of virus in my samples from Vietnam, Thailand, and Arizona. Even with large amounts of DNA, parvovirus DNA is very difficult to clone and as a result, cloning from the limited amount of native viral DNA was not achieved. I have been trying out as well as modifying several techniques during my Ph.D. and finally I was able to obtain the termini of several model densoviruses, like *Junonia coenia* densovirus (JcDNV) and *Pseudoplusia includens* densovirus (PiDNV), of which larger amounts were available. This should make it possible to obtain full-length PstDNV clones once enough material is available.

The sequence of PstDNV between the telomeres has been successfully amplified by PCR and its transcription mode could be studied and compared to *Aedes albopictus* densovirus (AalDNV), a virus that was tentatively classified in the same genus of *Brevidensovirus*. However, our results show that PstDNV should be classified in a new genus.

In 2009-2010, a serious outbreak caused by *Acheta domesticus* densovirus (AdDNV) led to a high mortality rate of crickets on many cricket farms in North America and Europe. We weekly received many cricket samples as well as aerosol specimens and other insects co-cultured in cricket facilities from North America, Europe, Japan, and China. I became involved in the AdDNV project with regard to diagnosis and geographical/phylogenetical analysis as well as on basic research on AdDNV. This also led to the discovery of three new cricket ssDNA viruses, i.e. two new densoviruses and one circular ssDNA virus.

In addition, my work on PstDNV-infected shrimps led to the discovery of two circular ssDNA viruses.

Summary

Single-stranded DNA viruses are among the smallest viruses, ranging from 17 to 25 nm in diameter. They include, among others, members of the *Parvoviridae* family with linear DNA genomes and the *Circoviridae* family with circular single-stranded DNA (ssDNA viruses). Many of these viruses have been found to be serious animal and human pathogens. In the past 40 years, numerous novel ssDNA viruses have been discovered and characterized from different hosts such as crustaceans, insects, and mammals. Recently, large numbers of new viruses and their isolates have been identified, especially with the advent of next-generation (deep) sequencing of environmental samples, revealing a rapid evolution rate and widespread occurrence of ssDNA viruses.

Many studies on the molecular biology, viral structure, and host interaction of ssDNA viruses are in progress. However, most of these studies are very challenging since these viruses generally lack a cell culture system for their *in vitro* propagation, particularly for parvoviruses.

The first objective of our current study focused on the transcription strategy of two tentative members of Brevidensovirus genus: *Penaeus stylirostris* densovirus (PstDNV) and *Aedes albopictus* densovirus (AalDNV). PstDNV was a monosense densovirus first identified in the 1980s. It caused significant, global economic losses during the last decades. Several incomplete PstDNV sequences deposited on Genbank were all derived from PCR amplification. However, those sequences still lack both palindromic extremities that are a hallmark of all parvovirus genomes. According to our experience, due to the low viral content of infected shrimps and the negative strandedness of the viral encapsidated DNA, obtaining infectious clones from native virus is extremely challenging. Therefore, complete PstDNV genome has not been achieved yet. During the past 30 years, despite efforts from scientists from all over the world and large investments, routinely useful primary cell cultures or cell lines of shrimp or other crustaceans (crab, lobster, and crayfish) have not been successfully developed. In this thesis, the expression strategy of PstDNV was successfully explored and confirmed via non-host cell line C6/36 and baculovirus system. As a comparison, we also studied the transcription map of *Aedes albopictus* densovirus (AalDNV), a chronic pathogen for *Aedes albopictus* larvae. These two viruses were classified into the same genus, *Brevidensovirus*, a rarely studied genus in the *Densovirinae* subfamily, *Parvoviridae* family. Our results showed quite different characteristics in their expression: PstDNV employed a

splicing mechanism for NS1 and two distant promoters (the NS2 promoter is located in the intron of NS1) for NS proteins instead of a leaky scanning and two overlapping promoters in AalDNV. This indicated that PstDNV should be classified in a separate genus.

The second objective was the geographycal analysis and the expression strategy of *Acheta domesticus* densovirus (AdDNV). This virus has been an epidemic pathogen for cultured house crickets in Europe but only became to a full-blown epidemic in Europe and North America in 2009-2010. AdDNV is an ambisense virus and was discovered more than 30 years ago. It was not further studied until recently, when devastating pandemics necessitated its characterization. Its complete genome has been cloned, sequenced and its unique expression strategy has been studied. AdDNV revealed unusual splicing features, suggesting a new genus *Pefudensovirus* within the ambisense densovirus group. Both NS and VP mRNA were spliced. Splicing in NS mRNA resulted in a set of three non-structural proteins NS1, NS2, and NS3. For the VP gene, there were two ORFs, ORF-A, which coded for VP2, and ORF-B. Connecting the two ORFs via splicing resulted in the removal of the N terminus of VP2 and three VP proteins, VP1, VP3, and VP4, were then produced. The N-terminal extension of VP1 contained a phospholipase A2 (PLA2) motif required to breach the endosomal membrane during cell entry. Subsequently, *Blattella germanica* densovirus (BgDNV) was shown to share similar characteristics in genome organization and gene expression.

Eight AdDNV isolates from different epidemics in Europe, Japan, and North America, from different hosts, *Acheta domesticus* and *Gryllodes sigillatus*, were cloned and sequenced. Phylogenetic analysis revealed that the virus responsible for the devastating outbreaks in 2009/2010 in Europe and North America were closely related and diverged around 2006. However, the virus responsible for the outbreak at the same time in Japan diverged probably much earlier suggesting other contributing factors in these pandemics.

As for other parvoviruses, the PstDNV genome should contain terminal hairpins essential for viral DNA replication by a rolling circle mechanism. It has been 40 years since the first report of this virus appeared, and its extremities are still a mystery. Cloning and sequencing of parvovirus ends are challenging tasks due to their highly stable hairpin structures, high GC content or genome packaging of minus rather than both strands. For the human B19 parvovirus, for example, the terminal sequences were solved only 19 years after it had been discovered and the main part of the genome had been sequenced. Our third objective was directed towards these problems. We developed a promising method, using

Pseudoplusia includens densovirus (PiDNV) and *Junonia coenia* densovirus (JcDNV) models, that could resolve the difficulties in the amplification and sequencing of the extremities. No hairpins were found by analyzing the extremities of PstDNV, but direct terminal repeats (DTRs) were found at both ends that could be essential in a novel replication model. Therefore, we propose a new classification for these two viruses.

In attempts to identify infectious agents, which caused high mortality rates in AdDNV-negative *Acheta domesticus* crickets, we discovered novel linear and circular ssDNA viruses. The fourth part of this thesis is focused on these new ssDNA viruses that infect different cricket species. We discovered two new densoviruses in *A. domesticus*, named *Acheta domesticus* mini ambisense densovirus (AdMADV) and *Acheta domesticus* segmented densovirus (AdSDNV). In addition, for the first time, a new circular DNA virus, named *Acheta domesticus* volvovirus (AdVVV), was discovered in different cricket species in North America and Japan. These viral genomes displayed no nucleotide sequence identity to any virus deposited in GenBank. AdMADV has a small ambisense genome size of 4945 nts and the related 199 nt-ITRs that fold into Y-shaped hairpin structure at both telomeres. Overall genome organization of this virus was found similar to that of other members of *Densovirus* genus. The second densovirus AdSDNV has an unusual property that made it differs from other typical densoviruses; the complete non-structural coding region of 3234 is flanked by two T-shaped telomeres. Because no VP gene was detected, we assume that there may be a VP- coding fragment separated from NSs.

First circular Rep-encoding ssDNA (CRESS-DNA) viruses (CVs) were also detected and cloned from PstDNV-infected *Penaeus monodon*. These viruses have small but diverse genome sizes, 1788 (Pm-associated CV or PmaCV-2) and 1777 (PmCV-1) nts containing two main ORFs. The Rep-coding ORF of PmCV-1 shared 30% identity at 90% coverage with that of cycloviruses and copepod *Labidocera aestiva* circovirus.

Our results confirmed the rapid and widespread evolution of this ssDNA virus group.

Résumé

Les virus à ADN simple brin figurent parmi les plus petits virus connus à ce jour, leur taille se situant entre 17 et 25 nm. Ils comprennent, entre autres, les membres de la famille des *Parvoviridae* dont le génome est un ADN linéaire et ceux de la famille des *Circovoridae* dont le génome est circulaire. Plusieurs de ces virus sont des agents de maladies, certaines graves, aussi bien chez l'homme que chez les animaux. Au cours des 4 dernières décennies, plusieurs nouveaux virus à ADN simple brin ont été isolés et caractérisés à partir de différents hôtes, depuis les crustacés et les insectes jusqu'aux mammifères. Récemment, l'avènement des nouvelles technologies de séquençage a permis l'identification d'un grand nombre de nouveaux virus et de leurs isolats à partir d'échantillons récoltés dans l'environnement et a révélé une très grande diversification et dispersion géographique des virus à ADN simple brin. De nombreuses études de biologie moléculaire et de structure des virions ainsi que l'analyse de leurs interactions avec leurs hôtes sont en cours pour caractériser ces virus. Toutefois, la plupart de ces études se heurtent au fait que l'on ne dispose pas de cultures cellulaires permettant leur multiplication, comme c'est le cas par exemple pour plusieurs parvovirus.

Le premier objectif de notre recherche a été de décrypter la stratégie de transcription de deux parvovirus d'invertébrés, le densovirus de la crevette *Penaeus stylirostris* (PstDNV) et celui du grillon *Acheta domesticus* (AdDNV). Ces deux virus sont responsables d'importantes épizooties dans les élevages intensifs de leurs hôtes. Concernant le PstDNV, il a été identifié dans les années 1980 et a rapidement causé, au cours des trois dernières décennies des pertes importantes dans plusieurs régions d'élevage intensif à grande échelle de crevettes. Plusieurs séquences incomplètes, toutes obtenues par PCR, ont été déposées dans GenBank. Ces séquences ont montré une organisation des régions codantes de type monosens, semblable à celle connue de certains densovirus de moustiques appartenant au genre *Brevidensovirus*. Toutes ces séquences sont caractérisées par l'absence d'extrémités palindromiques, ce qui est la règle chez tous les autres parvovirus. Par suite de la faible teneur en virus des échantillons dont nous disposions et le fait que les particules virales renferment essentiellement des brins de polarité négative, le clonage d'une séquence complète infectieuse d'ADN bicaténaire du génome du PstDNV est un véritable défi qui n'a pas encore été surmonté à ce jour. De même, les nombreux essais dans plusieurs laboratoires d'établir des cultures primaires ou une lignée cellulaire de crustacé pour

multiplier ce virus ont tous échoué jusqu'à présent. En dépit de ces difficultés nous avons réussi à décrypter la stratégie d'expression du PstDNV en transfectant la lignée cellulaire C6/36 de moustique et en utilisant un système d'expression baculovirus. Nous avons étudié en parallèle les modalités de transcription du densovirus d'*Aedes albopictus*, un virus pathogène des larves de ce diptère et qui infecte chroniquement certaines souches de la lignée C6/36. Ces deux virus ont été classés dans le même genre *Brevidensovirus* de la sous-famille des *Densovirinae.* Toutefois, nos résultats ont mis en évidence des différences significatives entre ces deux virus. Le PstDNV utilise un promoteur spécifique et un épissage pour l'expression de la protéine de réplication NS1 et un deuxième promoteur localisé dans l'intron de NS1 pour exprimer la protéine NS2. Par contre, l'expression des deux protéines NS de l'AalDNV est sous le contrôle de deux promoteurs se chevauchant et NS2 est traduit par un mécanisme de «leaky scanning». À notre avis, ces différences justifient la classification du PstDNV dans un nouveau genre.

Le deuxième objectif était l'analyse geographique et la stratégie de l'expression de l'*Acheta domesticus* densovirus (AdDNV). L'AdDNV a été découvert il y a plus de 30 ans mais n'a été caractérisé que récemment, suite à des épidémies foudroyantes dans les élevages en masse de son hôte en Amérique du Nord. Le génome complet de l'AdDNV a été cloné et séquencé et sa stratégie d'expression s'est révélée unique, suggérant qu'il correspond à un nouveau type au sein du groupe des densovirus ambisens. L'épissage différentiel des messagers NS génère 3 protéines non structurales NS1, NS2 et NS3. Les séquences codantes des protéines structurales sont portées par deux cadres de lecture (ORFs), l'ORF-A et l'ORF-B. L'ORF-A code la protéine majeure VP2 et, par un épissage éliminant la région N-terminale de VP2 et mettant en phase les deux ORFs, les protéines VP1, VP3 et VP4 sont produites. La séquence N-terminale de VP1 contient le motif phospholipase de type A2 requis pour dégrader la membrane endosomiale au moment de l'entrée du virus. Par la suite, le même type d'organisation génomique et d'expression a été démontré pour le densovirus BgDNV de blatte.

Nous avons réalisé le clonage et le séquençage complet de huit isolats d'AdDNV à partir de cadavres *d'Acheta domesticus* et de *Gryllus sigillatus* récoltés au cours de pandémies en Europe, au Canada, aux États-Unis et au Japon. Les analyses phylogénétiques ont montré que les virus responsables d'épizooties dévastatrices en 2009-2010 en Europe et en Amérique du Nord sont fortement apparentés et ont divergé vers 2006. Le virus responsable de l'épizootie survenue au Japon à la même époque a divergé

vraisemblablement il y a beaucoup plus longtemps, ce qui suggère que différents facteurs ont contribué à l'apparition de ces pandémies.

Comme celui des autres densovirus, le génome du PstDNV doit posséder à chaque extrémité une séquence palindromique nécessaire pour la réplication de l'ADN sur le modèle du cercle roulant (RCR). Ce virus est connu depuis plus de 40 ans et la structure des séquences terminales de son génome restent une énigme. Le clonage et le séquençage des extrémités des génomes de parvovirus se sont toujours avérés difficiles à cause de leur forte teneur en GC rendant très stables les structures terminales en épingle à cheveux et, également, du fait de l'encapsidation majoritaire de brins négatifs. Tel est le cas, par exemple pour celles du parvovirus humain B19 dont la séquence de ses extrémités n'a été obtenue que 19 ans après sa découverte. Notre troisième objectif a été de résoudre ces problèmes. La méthode que nous avons utilisée pour séquencer les extrémités du génome du PstDNV est la même que celle qui a été utilisée pour les extrémités des densovirus PiDNV et JcDNV dans notre laboratoire. A notre grande surprise, aucune structure en épingle à cheveux n'a été trouvée, mais seulement des répétitions directes aux deux extrémités qui permettent de proposer un nouveau modèle de réplication. Nous proposons donc une nouvelle classification pour ce virus.

L'absence d'AdDNV dans plusieurs échantillons de grillons provenant d'élevages à taux de mortalité élevé nous a conduit à rechercher et à découvrir de nouveaux types de virus à ADN simple brin linéaire ou circulaire. La quatrième partie de cette thèse est consacrée à la présentation de ces nouveaux types de virus infectant différentes espèces de grillons. Nous avons découvert deux nouveaux densovirus chez l'*Acheta domesticus* nommés; *Acheta domesticus* mini-ambisens densovirus (AdMADV) et *Acheta domesticus* segmented densovirus (AdSDNV). En outre, et pour la première fois, nous avons identifié un nouveau virus à ADN circulaire que nous avons appelé *Acheta domesticus* volvovirus (AdVVV) et il a été identifié à partir de grillons de différentes espèces d'Amérique du Nord et du Japon. Ces virus ne présentent aucune séquence d'homologie avec aucuns des génomes viraux déposés dans GenBank. L'AdMADV a un petit génome ambisens de 4945 nt et a des «Inverted terminal repeats» (ITRs) de 199 nt qui constituent une structure en épingle à cheveux en forme d'Y dans ses télomères. En général, l'organisation du génome de ce virus a été trouvée comparable à celle d'autres membres du genre *Densovirus*. Le deuxième densovirus; AdSDNV, possède une propriété unique qui n'a pas été retrouvée chez les autres Densovirus typiques; la région non-structurale complète de 3234 est couverte par les

deux télomères en forme de T. De plus, le gène VP n'y a pas été détecté. Nous supposons qu'il y a peut-être un fragment codant pour le VP qui est séparé du NS. Une autre possibilité qui peut être prise en compte est que ce virus pourrait utiliser la capside d'autres virus, probablement celle d'AdVVV, car, nous avons remarqué que les grillons infectés par l'AdSDNV étaient positifs à l'AdVVV par PCR. Des génomes viraux à ADN circulaire ont également été clonés et séquencés à partir de la crevette *Penaeus monodon*. Ces virus ont des génomes de tailles diverses allant de 1777 (PmCV-1) et 1788 (PmaCV-2) nts possédant au moins deux cadres de lecture. L'ORF Rep de PmCV-1 codant la protéine de réplication représente 90% d'identité et 30% de couverture avec les cyclovirus en général et le circovirus du copépode *Labidocera aestiva* en particulier. Ainsi, nos résultats confirment l'évolution rapide et l'expansion de ces groupes de virus à ADN simple brin chez les invertébrés.

List of figures

Introduction

Results

Discussion

Annexes

Annexe 1

Annexe 2

Annexe 10

List of tables

List of publications

1. The *Acheta domesticus* densovirus, isolated from the European house cricket, has evolved an expression strategy unique among parvoviruses. Liu K, Li Y, Jousset FX, Zadori Z, Szelei J, Yu Q, Pham HT, Lépine F, Bergoin M, Tijssen P. J Virol. 2011 Oct; 85(19):10069-78. doi: 10.1128/JVI.00625-11. Epub 2011 Jul 20. PMID; 21775445.

2. Billions and billions sold: Pet-feeder crickets (Orthoptera: Gryllidae), commercial cricket farms, an epizootic densovirus, and government regulations make for a potential disaster. David B. Weissman, David A. Gray, Hanh Thi Pham & Peter Tijssen. Zootaxa. 2012 Oct; 3504: 67–88.

3. *Pseudoplusia includens* densovirus genome organization and expression strategy. Huynh OT, Pham HT, Yu Q, Tijssen P. J Virol. 2012 Dec; 86(23):13127-8.

4. *Acheta domesticus* volvovirus, a novel single-stranded circular DNA virus of the house cricket. Pham HT, Bergoin M, Tijssen P. Genome Announc. 2013 Mar 14;1((Hsu *et al.*)2):e0007913. doi: 10.1128/genomeA.00079-13. PMID: 23516206.

5. New Volvovirus Isolates from *Acheta domesticus* (Japan) and *Gryllus assimilis* (United States). Pham HT, Iwao H, Bergoin M, Tijssen P. Genome Announc. 2013 Jun 27;1(3). pii: e00328-13. doi: 10.1128/genomeA.00328-13. PMID: 23792751.

6. Comparative genomic analysis of *Acheta domesticus* densovirus isolates from different outbreaks in Europe, North America, and Japan. Pham HT, Iwao H, Szelei J, Li Y, Liu K, Bergoin M, Tijssen P. Genome Announc. 2013 Aug 15;1(4). pii: e00629-13. doi: 10.1128/genomeA.00629-13. PMID: 23950131.

7. *Junonia coenia* densovirus (JcDNV) genome structure. Pham HT, Huynh OT, Jousset FX, Bergoin M, Tijssen P. Genome Announc. 2013 Aug 8; 1(4). pii: e00591-13. doi: 10.1128/genomeA.00591-13. PMID: 23929483.

8. Expression strategy of *Aedes albopictus* densovirus. Pham HT, Jousset FX, Perreault J, Shike H, Szelei J, Bergoin M, Tijssen P. J Virol. 2013 Sep; 87(17):9928-32. doi: 10.1128/JVI.01259-13. Epub 2013 Jun 26. PMID: 23804640.

9. A novel ambisense densovirus, *Acheta domesticus* mini ambidensovirus, from crickets. Pham HT, Yu Q, Bergoin M, Tijssen P. Genome Announc. 2013 Nov 7; 1(6). pii: e00914-13. doi: 10.1128/genomeA.00914-13. PMID:24201197.

10. A circo-like virus isolated from *Penaeus monodon* shrimps. Pham HT, Yu Q, Boisvert M, Van HT, Bergoin M, Tijssen P. Genome Announc. 2014 Jan 16; 2(1). pii: e01172-13. doi: 10.1128/genomeA.01172-13. PMID: 24435870.

11. Complete sequence and expression strategy of *Penaeus stylirostris* densovirus (PstDNV). Hanh T. Pham, Françoise-Xavière Jousset, Jonathan Perreault, Hiroko Shike, Jozsef Szelei, Max Bergoin, Peter Tijssen. J Gen. Virol. (Manuscript, to be resubmitted in J Gen Virol. 2014).

12. A brevidenso-like virus, *Acheta domesticus* segmented densovirus AdSDNV, isolated from diseased crickets in US and Canada. Pham HT *et al.*, in preparation.

List of abbreviations and symbols

3'-OH	3'-hydroxyl group
3D	Three-dimensional
aa	Amino acid
AaeDNV	*Aedes aegypti* densovirus
AalDNV	*Aedes albopictus* densovirus
AAV	Adeno-associated virus
AdDNV	*Acheta domesticus* densovirus
AdMADV	*Acheta domesticus* mini ambidensovirus
AdSDNV	*Acheta domesticus* segmented densovirus
AdVVV	*Acheta domesticus* volvovirus
AMDP	Aleutian mink disease parvovirus
APARs	Autonomous parvovirus-associated replication bodies
APC/C	Cellular anaphase promoting complex/cyclosome
ATF/CREB	Activating transcription factors/ cAMP response element binding
AVG2	Avian gyrovirus 2
B19	Human parvovirus B19
BBTV	Banana bunchy top virus
BgDNV	*Blattella germanica* densovirus
BmDNV	*Bymbyx mori* densovirus
Cap	Capsid protein
CAT	Chloramphenicol acetyltranferase
CAV	Chicken infectious anemia virus
CFS	Cerebrospinal fluid

CNS	Central nervous system
CpDNV	*Culex pipiens* densovirus
CppDNV	*Culex pipiens pallens* densovirus
CPV	Canine parvovirus
CREB-BP1	Cyclic AMP-response element binding protein
CRESS	Circular Rep-encoding ssDNA
Crf	Covalently replicative-intermediate DNA form
CRM1	Chromosomal region maintenance 1
CSS-DNA	Circular single-stranded DNA
CyVs	Cycloviruses
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
dNTP	Deoxyribonucleotide triphosphate
DNVs	Densoviruses
DpDNV	*Dendrolimus punctatus* densovirus
DPE	Downstream promoter element
DpIDNV	*Dysaphis plantaginea* densovirus
dsDNA	Double-stranded DNA
DsDNV	*Diatraea saccharalis* densovirus
DSP	Dual specificity protein phosphatase
DTR	Direct terminal repeat
EDTA	Ethylene diamine tetraacetic acid
EM	Electron microscopy
FPV	Feline panleucopenia virus
GEMB	Glucocorticoid modulating element-binding protein
GmDNV	*Galleria mellonella* densovirus

HaDNV	*Helicoverpa armigera* densovirus
HeDNV	*Haemagogus equines* densovirus
HePV	Hepatopancreatic parvovirus
HIV	Human immunodeficiency virus
IHHNV	Infectious Hypodermal and Hematopoietic Necrosis parvovirus
Inr-box	Transcription initiation box
ITR	Inverted terminal repeat
JcDNV	*Junonia coenia* densovirus
kb	Kilobase
kDa	KiloDalton
MCM3	Mini-chromosome maintenance protein 3
MCMS	Mid-crop mortality syndrome
MlDNV	*Mythimna loreyi* densovirus
MpDNV	*Myzus persicae* densovirus
Mot1	Modifier of transcription 1
mRF	Monomer of replicative-intermediate DNA form
MSV	Maize streak virus
MVM	Minute Virus of Mice
NC2	Negative cofactor 2
NES	Nuclear export signal
NF-κB	Nuclear factor kappa-light-chain-enhancer of activated B cells
NLS	Nuclear localization signals
nm	Nanometer
NS	Non-structural protein
NTP	Nucleotide triphosphate
Oc8	Octanucleotide motif

ORF	Open reading frame
OriL	Viral origin in the left telomere sequence
OriR	Viral origin in the right telomere sequence
p.i.	Post infection
p.t.	Post transfection
PcDNV	*Planococcus citri* densovirus
PCNA	Proliferating cell nuclear antigen
PCR	Polymerase chain reaction
PCV	Porcine circovirus
PfDNV	*Periplaneta fuliginosa* densovirus
PiDNV	*Pseudoplusia includens* densovirus
PIF	Parvovirus initiation factor
PL	Postlarvae
PLA_2	Phospholipase A2
PmCV-1	*Penaeus monodon* circo-like virus 1
PmaCV-2	*Penaeus monodon* associated-circovirus 2
PmDNV	*Penaeus monodon* densovirus
PMWS	Postweaning multisystemic wasting syndrome
poly(A)	Polyadenylation
Polδ	DNA polymerase δ
PpDNV	*Papilio polyxenes* densovirus
PRE	Polycomb response elements
PstDNV	*Penaeus stylirostris* densovirus
RACE	Rapid amplification of cDNA ends
RCR	Rolling circle replication
Rep	Replication protein

RF	Replicative-intermediate DNA form
RFC	Replication factor C
RHR	Rolling hairpin replication
RPA	Single-strand binding replication protein A
RPAs	RNase Protection Assays
SF3	Superfamily 3
SfDNV	*Sibine fusca* densovirus
Smn	Survival motor neuron protein
SMV	Spawner-isolated mortality virus
Sox	Sry-related HMG box
ssDNA	Single-stranded DNA
TTV	Torque teno virus
UTR	Untranslated region
VP	Structural protein
VP1up	Unique region of VP1
VLPs	Virus-like particles

1.Introduction

1.1. Industrial shrimp and cricket farming and epidemic caused by parvoviruses

1.1.1. Shrimp farming and viral epidemic disease

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. Along with the rapid expansion of the shrimp industry, the uncontrolled growth of these monocultures, favoured epizootic diseases of viral or bacterial etiology causing important economic losses. Viral diseases are the most hazardous impediments to shrimp culture (Flegel, 2012). Although these shrimp viruses pose no threat to human health, they have been causing severe economic impacts throughout Asia and South America.

Viruses that have been recognized as causative agents of lethal diseases in shrimps belong essentially to families *Parvoviridae*, *Baculoviridae*, *Picornaviridae*, and Toga-like viruses. Some newly identified virus families have also been reported (Lightner *et al.*, 1997, Walker *et al.*, 2010). In the dense confined populations of shrimp ponds, virus infections spread rapidly and may wipe out the whole shrimp population. The uncontrolled introduction of shrimps from one area to another is a major risk factor of virus epidemics.

The major single-stranded DNA (ssDNA) viruses that caused significant diseases of farmed and wild shrimp are *Penaeus stylirostris* densovirus (PstDNV), previously known as Infectious Hypodermal and Hematopoietic Necrosis parvovirus (IHHNV), *Penaeus monodon* densovirus (PmDNV) or Hepatopancreatic parvovirus (HePV) and Spawner-isolated mortality virus (SMV), all members of the *Parvoviridae* family.

***Penaeus stylirostris* densovirus - PstDNV**

PstDNV was found to be the most virulent to farmed shrimp in the Americas during the 1990's. During this period, the estimated lost revenue due to PstDNV was about 0.5-1 billion USD (Lightner, 2005). It was first discovered in blue shrimp *Penaeus stylirostris* cultured in Hawaii in 1981, causing acute epizootics and up to 90% mass mortality in *P. stylirostris*, especially in juveniles and subadults (Lightner *et al.*, 1983). Shortly after, the virus was detected in farmed *P. vannamei* in the Americas and Pacific coast, Southeast USA, Mexico, Ecuador, Peru, Brazil, Caribbean, Central America, Hawaii, Guam, Tahiti, New Caledonia, and in *P. monodon* in Southeast Asia, Singapore, Malaysia, Thailand, Indonesia, Taiwan, Philippines and

Vietnam (Flegel, 1997, Lightner, 1996a, Lightner, 1996b). It is a cosmopolitan virus, reported from natural infections of *P. stylirostris, P. vannamei, P. occidentalis, P. californiensis, P. monotone, P. semisulcatus, and P. japonicus.* The PstDNV-experimental infection has also been confirmed in *P. setiferus, P. dourarum, P. aztecus*, and *P. indicus*. Only *P. merguiensis* appears to be resistant to PstDNV (Flegel, 1997, Lightner, 1999). The isolated virus from the Philippines was shown to be virtually identical to PstDNV that was first detected in Hawaii (Lightner, 1999). It shares a high sequence identity to the Hawaiian with only 0.2% nucleotide sequence difference from a 2.9 kb fragment of the viral genome (Tang *et al.*, 2003). As of 2003, the US is the only country that can claim to have PstDNV-free zones. This was achieved with the development and use of SPF specific pathogen-free (SPF) shrimp stocks. Recently, since July 2004, high mortalities are frequently encountered in postlarvae (PL) and subadults of giant freshwater shrimps, *Macrobrachium rosenbergii*, in Taiwan and Malaysia. By pathological examination and molecular assays, PstDNV seems to be involved in these outbreaks (Chia et al., 2006, Hazreen Nita et al., 2012). In contrast to *P. stylirostris* and *Macrobrachium rosenbergii* hosts, PstDNV is less virulent in *P. vannamei* and *P. monodon*. Vertically infected early larvae and PL are more severe but do not become diseased before the age of 35 days and within the size range 0.05 to 1 g (juvenile life). Infected adults seldom show signs of the disease or mortality. For that reason, most of the farmers underestimate its danger to cultured shrimp. The individuals of the populations that survive the PstDNV epizootics can carry virus for life and pass it onto others by vertical transmission. *P. monodon* in Africa and *Cherax quadricarinatus* in Australia have been found to contain PstDNV-related sequences in their chromosomes (Rusaini *et al.*, 2013, Tang *et al.*, 2006).

***Penaeus monodon* densovirus - PmDNV**

Similar to PstDNV, PmDNV causes a widespread disease in PL, juveniles and adults. PmDNV was first reported in Singapore in *P. merguiensis* and *P. indicus* (Chong *et al.*, 1984), in *P. chinensis (*Lightner *et al.*, 1985*)*, and in *P. monodon* (Sukhumsirichart *et al.*, 1999). Several penaeid shrimp PmDNV infections were also described in Asian, Australian, American and African regions (Bonami *et al.*, 1995, Flegel, 2006, Lightner, 1996b). Freshwater shrimp *Macrobrachium rosenbergii* was also found to be infected with PmDNV (Anderson *et al.*, 1990, Gangnonngiw *et al.*, 2009). An experimental model of PmDNV horizontal transmission was reported in *P. monodon* PL by feeding PL with infected PL or Artemia (Catap *et al.*, 2005, Sivakumar *et al.*, 2009). Evidence exists that PmDNV is transmitted vertically from broodstock to progeny and horizontally during the PL stages (Brock *et al.*, 1990, Manivannan *et al.*, 2002).

However, in studies on captured Thai broodstock specimens (Flegel, 1997), none were found to show the characteristic histopathology of PmDNV. This suggests that the virus may not originate with the broodstock but with artemia or other carriers present in the cultivation system. Moreover, the cannibalism of shrimp populations causes horizontal transmission of this virus and is a serious problem for shrimp farmers.

Spawner-isolated mortality virus - SMV

Spawner-isolated mortality virus (SMV) was first identified in *P.monodon* in Australia (Fraser *et al.*, 1996). From 1994 to 1996, SMV was one of several viruses associated with mid-crop mortality syndrome (MCMS), which caused significant mortalities of broodstock, juveniles and subadults of *P. monodon* cultured in Australia (Owens *et al.*, 1998). In Philippines, *P. monodon* infected with luminous vibriosis (*Vibrio harveyi)* were also found to be infected with SMV (Albaladejo *et al.*, 1998). Infection and disease due to SMV have only been reported from cultured or captive wild adult *P. monodon* and cultured *Cherax quadricarinatus* (Owens *et al.*, 2000). Experimental infections have also resulted in mortalities in *P. esculentus, P. japonicus, and P. merguiensis* and *Metapeneus ensis.* There are no diagnostic gross signs for SMV disease. Juvenile prawns in grow-out ponds with clinical viral infections may exhibit signs such as discoloration, redness of the carapace and pleopods, lethargy, fouling and anorexia. Red feces are also a characteristic of this disease (Fraser *et al.*, 1996).

1.1.2. Cricket industry and epidemic diseases caused by *Acheta domesticus* densovirus (AdDNV)

Raising house crickets (*A. domesticus)* has been developed to a large-scale industry (Figure 1-1) in Europe and North America since the 1940s (Weissman *et al.*, 2012). This industry supplies billions of live crickets every year as fish baits, food for pet reptiles, zoo amphibians and other purposes. In the pharmaceutical industry, crickets, instead of the American cockroach, are being used widely for physiological and toxicological studies. China, Thailand, and Japan have been raising crickets as house pets or daily food. According to an American Pet Products Association National Pet Owners Survey, reptiles are becoming common pets in many US households. In 2013, about 11.5 million reptiles were being kept as pets in US and crickets are staple, tender protein source for them. Cricket and reptile businesses hence have been growing in parallel and their business values were estimated at about 650 million and 4.5 billion in 2010, respectively. The biggest cricket supply is from US; major facilities ship 5.5 million crickets weekly, mainly for the pet industry (Szelei *et al.*, 2011,

Weissman *et al.*, 2012). Many cricket breeders also raise other insects, such as superworms (*Zophobas morio*), waxworms (*Galleria mellonella*), or mealworms (*Tenebrio molitor*).

Crickets are very easy to rear in small or big scale. However, due to the dense populations, infectious diseases can spread out quickly. AdDNV is 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, only one small outbreak related to this virus in 1988 was reported from the Southeastern US (Styer *et al.*, 1991). Later, in 2009, this virus has suddenly devastated North American cricket operations. Heavy mortality rates of cultured crickets were reported, up to 80-100% from California, Florida, Alberta, and Québec (Szelei *et al.*, 2011). During this time, it was almost impossible to get free-AdDNV crickets. Several cricket farms declared bankruptcy, risking to ruin the 4.5 billion dollar-pet industry in the US and Canada. The producers have been trying to improve the survival of crickets with the selection of AdDNV-resistants through several generations but they did not succeed. Some other growers switched to other species, such as the *Gryllus assmilis* (Jamaican field cricket), to avoid severe losses. However, these species have certain diapause stages during their life and therefore they are not reproductive enough to supply the commercial demands. *A. domesticus* does not need such resting periods but it is the most susceptible host for AdDNV. *G. assimilis* was known resistant to AdDNV but can contain virus confirmed by Polymerase chain reaction (PCR) (Szelei *et al.*, 2011).

Figure 1-1. The set-up of breeding containers in cricket farms

AdDNV was detected by electron microscopy (EM) and PCR in cricket droppings, air filters from the furnace of infected farms. In addition, virus are also present in "superworms" (*Zophobas morio*), mealworms (*Tenebrio molitor*) and "waxworms" (*Galleria mellonella*) that have been raised together in cricket farms (Szelei *et al.*, 2011). This suggested that besides fecal-oral and aerosol transmission, many insects can be virus-carrier and should be considered among the vectors for AdDNV-mechanical transmission.

AdDNV is a pathogenic virus to crickets, especially in the last larval stage and young adults. Infected crickets become sluggish, cannot jump high, and usually lie on their back, display anorexia and die slowly within 10-14 days. Infected females lay much fewer eggs than healthy ones. Histological analysis of AdDNV positive crickets showed viral infection in the epidermal tissue, tracheal and midgut epithelial cells, fat body and Malpighian (Styer *et al.*, 1991, Szelei *et al.*, 2011). Several AdDNV strains related to different reported outbreaks in Europe and US have been cloned and sequenced (Szelei *et al.*, 2011). Sequence analysis revealed that AdDNV might have existed and circulated in North America several years before the outbreak. The 1977-European and 1988-American isolates are the same. Meanwhile, the 2009/2010-American isolate diverged in 2006 from the European strain and was responsible for the viral epidemic at that time in US and Canada (Szelei *et al.*, 2011).

1.2. Literature of molecular biology of densovirus

1.2.1. Classification and genome organization

Densoviruses (DNVs) are autonomous parvoviruses that infect a wide range of invertebrates and are often related to pathogenic diseases of their hosts. They belong to the subfamily *Densovirinae* in the *Parvoviridae* family. In contrast, members in the *Parvovirinae* subfamily have vertebrate hosts (Figure 1-2). Currently, phylogenetic analysis has been an important tool for clustering viruses. However, features such as genome organization, genome size and secondary structures of their terminal sequences, a distinctive hallmark of the parvovirus family, are also critical criteria for virus classification.

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. According to the newest report from the International Committee on Taxonomy of viruses (ICTV), the *Densovirinae* subfamily is divided into at least four genera: *Densovirus*, *Pefudensovirus*,

Iteravirus and *Brevidensovirus* (Tijssen *et al.*, 2011). Members of the genera *Densovirus* and *Pefudensovirus* possess ambisense genomes i.e. non-structural (NS) proteins coding sequences are located in the 5' half on one strand, whereas the structural (VP) proteins coding sequences are located in the other half on the complementary strand. Members of the *Iteravirus* and *Brevidensovirus* genera have a monosense gene organization, the regions coding for NS and VP being located in the 5' and 3' halves, respectively.

1.2.1.1. *Densovirus* genus

Members of this genus are defined by their ambisense genome, about 6 kilobase (kb) in size and by having the longest inverted terminal repeats (ITRs), over 500 nucleotides (nts), among parvoviruses. *Galleria mellonella* densovirus (GmDNV) was the first densovirus isolated from insects (Meynardier *et al.*, 1964). Infectious clones were obtained and sequenced showing that both long ITRs of GmDNV are identical, 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). This terminal structure is very similar to that of Adeno-associated viruses (AAVs). Minus and plus strands package in separate virions in a 1/1 ratio. Three open reading frames (ORFs) code for NS proteins under only one transcriptional promoter on the 5' half of one strand. The largest ORF codes for NS1 (Rep protein) and contains the NTPase motif that is typically found in NS1 of other parvoviruses (Figure 1-3). NS2 completely overlaps NS1 and its start codon is only 5 nts downstream of that of NS1. The left most ORF coding for NS3 is in frame with and separated from NS1 only by a TAA stop codon. VP gene containing PLA2 motif is on the right part of the viral genome. Other members of the *Densovirus* genus include the *Junonia coenia* densovirus (JcDNV) isolated from the common buckeye butterfly (Dumas *et al.*, 1992, Jourdan *et al.*, 1990), *Diatraea saccharalis* densovirus (DsDNV) isolated from the sugar cane borer larvae from Brazil (Kouassi *et al.*, 2007), and *Mythimna loreyi* densovirus (MlDNV) isolated from the noctuid maize worm in Egypt. *Helicoverpa armigera* densovirus (HaDNV) and *Pseudoplusia includens* densovirus (PiDNV) isolated from cotton bollworms and soybean loopers are two new members of this genus (El-Far *et al.*, 2012, Huynh *et al.*, 2012).

Only JcDNV and MlDNV have a wide host range including several noctuid pests such as *Spodoptera littoralis, Pecinophora gossypiella, Sesamia cretica, Chilo agamemnon* and *Ostrinia nubilalis* (Fediere *et al.*, 2004). Most densoviruses are potential agents for biological pest control and evidence proved that densoviruses do not infect vertebrate cells (El-Far *et al.*, 2004).

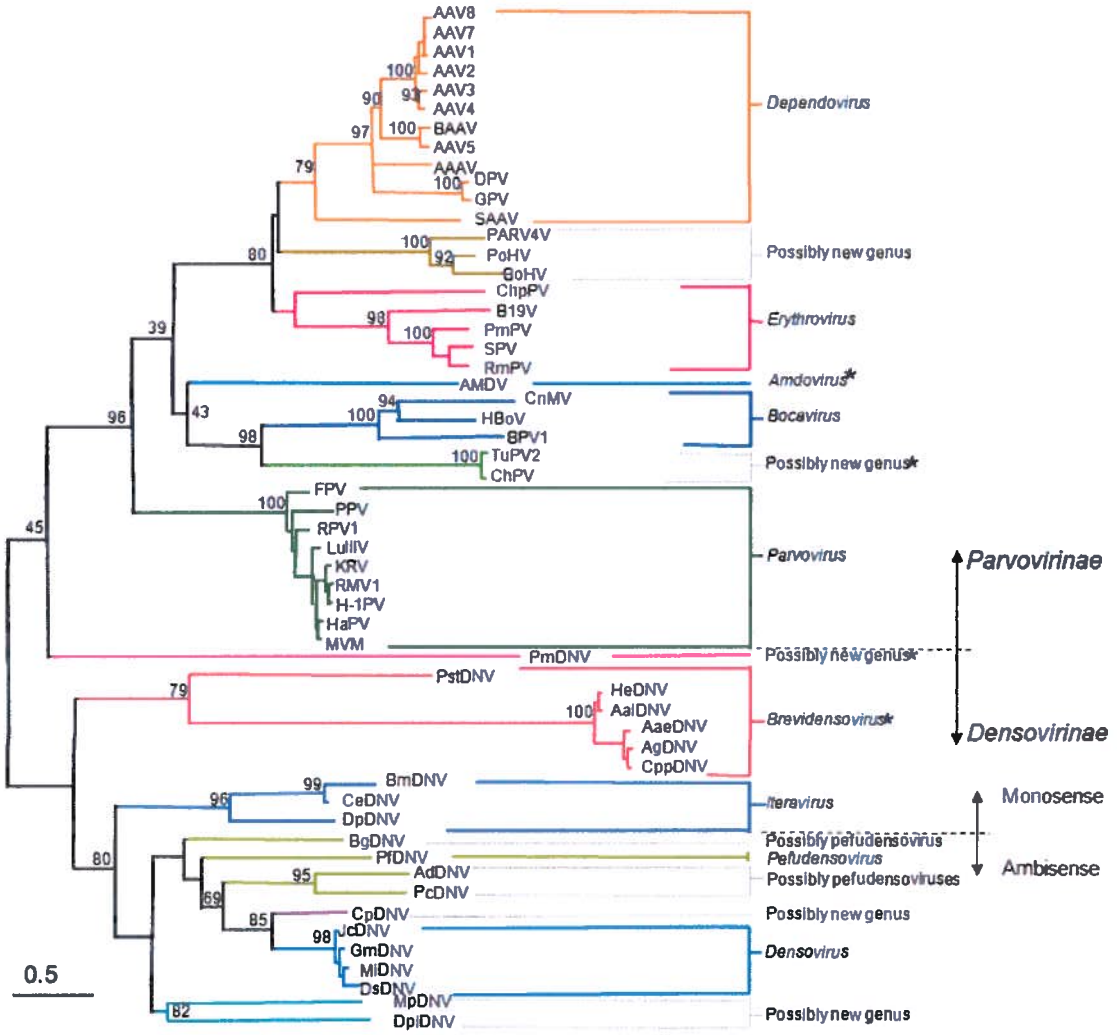


Figure 1-2. Phylogeny of members in *Parvovirinae* and *Densovirinea* subfamilies based on the NS1 protein sequences. The tree was constructed with the programs included in the Phylip package at the http://mobyle.pasteur.fr/cgi-bin/portal.py website (ClustalW-multialign, Phylip distance matrix, PROTDIST, Neighbor-Joining method, and phylogenetic tree drawing). The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) is shown next to the branches. Scale bar represents the rate of amino acid substitutions. The obtained tree was redrawn with like colours for viruses with similar genome organizations. This phylogenetic analysis distinguished the two subfamilies, as well as the monosense and ambisense densoviruses, and recognized the different genera as separate clades. Several other clades that are possible new genera, as described in the text, were also recognized. The asterix indicates clades of viruses that do not contain the phospholipase A2 motif in their capsid proteins (Tijssen *et al.*, 2011).

1.2.1.2. *Pefudensovirus* genus

Periplaneta fuliginosa densovirus (PfDNV) is the typical member of this genus. It was first isolated from smoky-brown cockroaches *Periplaneta fuliginosa* from a laboratory colony in Nagoya, Japan (Suto *et al.*, 1979). Later, in 1994, PfDNV was also reported in China and its genome was cloned and sequenced (Guo *et al.*, 2000, Hu *et al.*, 1994). Other putative members are AdDNV, *Blattella germanica* densovirus (BgDNV) and *Planococcus citri* densovirus (PcDNV). PfDNV and AdDNV are the first densoviruses found in *Orthoptera* and *Dictyoptera* species, respectively. As discussed above, AdDNV was detected in many cricket species. Several AdDNV isolates from different outbreaks in Europe and United Stated were cloned and sequenced (Szelei *et al.*, 2011). BgDNV was discovered from natural *Blatella germanica* cockroaches in a pig farm in North Carolina (Mukha *et al.*, 2006). The pefudensovirus genome is ambisense and about 5.5 kb in size including two imperfect palindromes of 125-175 nts at the two ends. The organization of pefudensovirus genomes is similar to that of members of the *Densovirus* genus but differ in having split ORFs for VPs. Three ORFs code for NS1, NS2, and NS3 on one strand and the two ORFs for VPs on the complementary strand (Figure 1-3). The 5'- and 3'- terminal sequences anneal by folding into an I-shaped hairpin. This hairpin configuration is similar to that of human virus B19 (Guo *et al.*, 2000, Mukha *et al.*, 2006, Yamagishi *et al.*, 1999). The position of PLA2 motif that previously identified on the N-terminal VP of porcine parvovirus (Zadori *et al.*, 2001), is located on the C-terminal of the small VP-ORF (Figure 1-3). In this genus, PcDNV genome still lacks the ITR sequences and possibly contains many indels that need to be re-sequenced.

1.2.1.3. *Iteravirus* genus

Members of this genus have monosense DNA genomes of about 5 kb. The ssDNA genome is packaged in virions with equal ratio of plus and minus strands. So far, all iteravirus genomes contain one ORF coding for four VP proteins and two ORFs coding for NS proteins. Iteravirus genomes possess two similar palindromic sequences of about 160 nts forming a J-shaped hairpin at each end (Bergeron *et al.*, 1993). The whole ITR sequence is about 230-270 nts. The typical member of this genus is *Bymbyx mori* densovirus (BmDNV-1). First reports on BmDNV-1 showed a complete genome with both ITRs (Bando *et al.*, 1990, Bando *et al.*, 1987) but this sequence contained several indels causing an inappropriate ORF map. Later, new clones of BmDNV were obtained and the correct sequence (Li *et al.*, 2001) shared up to 85,7%

identity to that of CeDNV, another iteravirus infecting *Casphalia extranea* (Fediere *et al.*, 2002). Many new viruses were recently identified in this genus such as, *Dendrolimus punctatus* densovirus (DpDNV) (Wang *et al.*, 2005), *Sibine fusca* densovirus (SfDNV) (Yu *et al.*, 2012a) and *Papilio polyxenes* densovirus (Yu *et al.*, 2012b). Iteraviruses are highly identical to each other at 75-85% nucleotide level. The hairpins are also conserved between them. BmDNV-1 and CeDNV ITRs differ only in one nt in their flip/flop sequence. Meanwhile, the ITR sequence of BmDNV-1 shared 98%, 80%, 37% identity to that of SfDNV PpDNV and DpDNV, respectively. Recently, a second densovirus HaDNV-1 was reported in *Helicoverpa armigera* which was clustered into this genus by phylogenic analysis but complete ITRs sequences has not been obtained (Xu *et al.*, 2012). Overall nucleotide sequence of HaDNV-1 reveals a highest similarity of 44.8% to that of DpDNV. The common pavoviral PLA2 motif is also present on VP1up region of all iteravirus genomes. Iteraviruses were found to infect the columnar tissue of midgut epithelium but other densoviruses do not (Tijssen *et al.*, 1995a). Unlike iteravirus, members of *Densovirus* genus such as JcDNV and GmDNV can not replicate in midgut cells. Therefore, these viruses need to across the midgut to reach the intestinal tissue. In JcDNV, this capacity was found to depend on the four aa residues A123, S167 and I174 and T177 that are near the 5-fold axis protrusions on major capsid protein VP4 since the replacing of these residues to those of GmDNV (A123V, S167T and I174V/T177D) caused significant decreases of virus transportation across the midgut of S. frugiperda (Multeau *et al.*, 2012). It might explain why JcDNV can infect a wide host range of lepidoptera in contrast to GmDNV.

1.2.1.4. *Brevidensovirus* genus

The monosense brevidensovirus genomes are about 4 kb and that has been known as the smallest genome among parvoviruses (Figure 1-3). This genus has some distinct characteristics from other viruses in the *Densovirinae* subfamily: no enzymatic core of PLA2 in their VP, three ORFs for NS and one for VP genes, 85% of encapsidated DNA have negative strand polarity and no ITRs but distinct palindromic sequences. Brevidensoviruses have short, dissimilar hairpin telomeres with a T-shaped structure. 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), *Haemagogus equines* densovirus - HeDNV (Paterson *et al.*, 2005). Their hosts are known as the epidemiologically essential vectors for several dangerous

pathogens to human such as yellow fever virus, dengue fever virus, malaria parasite, chikungunya, and West Nile virus. Therefore, mosquito brevidensoviruses are potential biological factors to control the paratransgenic pathogens. PstDNV is the only member in this group that has a crustacean host outside the insects. PstDNV was discovered over 30 years ago. Non-structural aa sequences of PstDNV and AaeDNV/AalDNV share highest identity to each other at about 30% (Shike *et al.*, 2000). Many studies on its pathogenicity, transmission, as well as different isolates reflecting geographic distribution have been reported. However, terminal sequences of the PstDNV genome are still unresolved. As for iteraviruses, little is known about the expression strategy of brevidensovirus.

1.2.2. Unclassified members of *Densovirinae* family

In addition to classified DNVs, there are numerous unclassified members. These viruses do not share any significant characteristics in their genomic sequence or organization to each other and to other DNVs in order to be classified into a defined genus (Tijssen *et al.*, 2011). So far, this suggests at least three new genera. The viruses in the first group, proposed as *Hepadensovirus* genus, have a 6.3 kb genome and comprise viruses from different shrimp hosts *Penaeus chinensis* densovirus (Bonami *et al.*, 1995), *Penaeus merguiensis* densovirus (La Fauce *et al.*, 2007) and *Penaeus monodon* densovirus (PmDNV) (Sukhumsirichart *et al.*, 2006). Thier genome is ambisense and does not contain ITRs or PLA2 motif-related sequences. Sequence analysis and phylogeny studies suggested that *Dysaphis plantaginea* densovirus - DpIDNV (Ryabov *et al.*, 2009) and *Myzus persicae* densovirus (MpDNV) (van Munster *et al.*, 2003) can be grouped together and separated from other densoviruses. They may belong to a new proposed genus *Aphidensovirus*. It was found that DpIDNV caused a negative impact on rosy aphid reproduction but virus infection could also induce winged morph production during parthenogenesis when they get DpIDNV infection (Ryabov *et al.*, 2009) to promote aphid spreading. *Culex pipiens* densovirus (CpDNV) is a 6 kb virus that has a similar ORF organization as members of the *Densovirus* genus. However, it has shorter ITRs with J-shaped form at the termini and lacks the (GAC) repeats that are involved as binding site of NS-1 in rolling-circle DNA replication. The NS1 and NS2 coding sequences were split into two ORFs. So far, this feature is unique among parvoviruses. The transcription profile of CpDNV was also studied; two distant promoters and splicing control the expression of NS1 and NS2 (Baquerizo-Audiot *et al.*, 2009). Those special characteristics suggest that CpDNV is a single member in a new genus *Cupidensovirus* (Tijssen *et al.*, 2011).

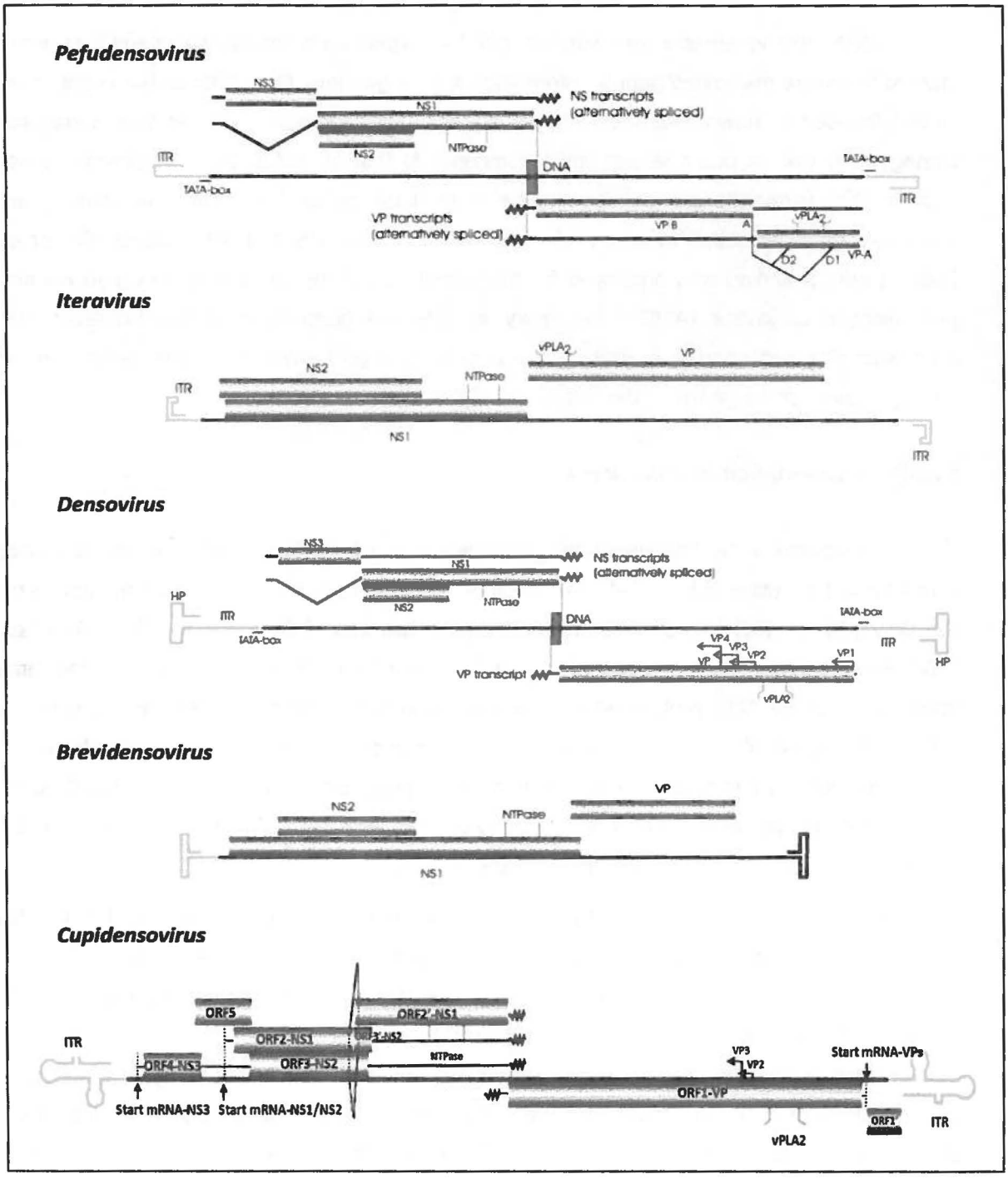


Figure 1-3. Genome organization, terminal structure, NTPase motif on NS protein, PLA2 motif on VP protein (except for Brevidensovirus) of different genera in *Densovirinae* subfamily (Tijssen *et al.*, 2006a).

1.2.3. Densovirus transcription strategy and regulation of expression

DNVs, like vertebrate parvoviruses, need to exploit an expression strategy that allows them to maximize the limited genetic information in their genome. Over the last ten years, many studies focused on how these small viruses express their genomes. However, their expression strategy is known in much lesser detail compared to that of vertebrate counterparts. It was showed that vertebrate parvoviruses utilize a complex pattern of alternative splicing and alternative polyadenylation to control the expression of their NS and VP proteins (Qiu *et al.*, 2006). Leaky scanning was employed for the expression of VP proteins in AAV and Aleutian mink disease parvovirus (AMDV) but rarely in other vertebrate parvoviruses. However, both leaky scanning and alternative splicing are commonly used by DNVs to increase the protein coding capacity of their genes. (Becerra *et al.*, 1985, Qiu *et al.*, 2007)

1.2.3.1. Transcription of *Densovirus*

Densovirus is the only genus with members of which the expression strategy has been characterized in details (Figure 1-4). Two sizes of NS transcripts were detected that start 23 nts (GmDNV), 27 nts (MIDNV) and 32 nts (JcDNV) downstream of the left ITRs (Abd-Alla *et al.*, 2004, Fediere *et al.*, 2004, Tijssen *et al.*, 2003). The NS3 ORF near the 5'-end of the large transcript yields the NS3 protein, whereas in the smaller transcript this NS3-ORF is deleted by splicing. Therefore, the context of the AUG of the first large ORF, NS1, in the smaller transcript is rendered poor by this splicing leading to additional translation of a second ORF, NS2 (Fediere *et al.*, 2004, Tijssen *et al.*, 2003). Only one promoter, with the TATA-box located within the upstream ITR, drives the expression of NS transcripts.

There is one large ORF4 on the right-hand half of the complementary strand that codes for about four VP proteins (Simpson *et al.*, 2002, Tijssen *et al.*, 1981). The generation of only one VP-mRNA suggested that leaky scanning was also employed for the expression of VP proteins (Tijssen *et al.*, 2003). The TATA-box of the promoter controlling VP transcription was also situated within right ITR sequence and produced one unspliced VP transcript that started at 25 nt (MIDNV) or 23 nt (GmDNV) upstream of the right ITR. Both initiation canonical sequences of transcription initiation box (Inr-box) of VP and NS transcripts in GmDNV and MIDNV are well-matched to the consensus motif G/A/T-C/T-A-T/G-T found for numerous *Drosophila* promoters (Purnell *et al.*, 1994). In JcDNV, two putative promoters P3 and P9 probably comprised the same crucial elements including TATA- and CAGT- boxes that are typical for insect promoters.

Activity of capsid promoter P9 of JcDNV was reported in insect (cells) *in vitro* and *in vivo*. Fusion of the lacZ gene within the VP gene resulted in a measurable expression of beta-galactosidase (Giraud *et al.*, 1992, Royer *et al.*, 2001). NS and VP promoters of GmDNV could drive the expression of the chloramphenicol acetyltranferase (CAT) gene in larvae (Tal *et al.*, 1993).

Densovirus members possess 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 overlapped with the NS stop codon. On the other hand, the poly(A) of the VP transcript completely overlapped the NS coding sequence and was 58 nts downstream of the VP stop codon. Therefore, the VP and NS transcripts have 56- (MIDNV) and 58- (GmDNV) nt overlapping sequences. Furthermore, the 5'-untranslated region (UTR) of NS transcripts is rather long, about 90 nts, whereas it is only 7 nts in the VP transcript to promote leaky scanning of the latter. Recently, it was reported that the deletion of a 557-bp fragment between SpeI (nt 3075) and BstZ17I (nt 2547) containing the overlapping region of NS and VP transcripts in JcDNV genome could diminish the activity of the P9 capsid promoter of JcDNV to undetectable levels. In contrast, the P9 activity was recovered when this fragment was reintroduced, even in the reversed direction, suggesting the presence of an enhancer or a *cis*-acting element that affects the expression activity of P9 (Shirk *et al.*, 2007). Indeed, three putative polycomb response elements (PRE) were predicted in this region within 300 nts around the poly(A) signals of NS and VP. Multiple putative PRE, Sry-related HMG box (Sox) or cyclic AMP-response element-binding protein (CREB-BP1) binding sites were also found to be distributed within the ITRs, 5'-NS or 3'-VP gene (Shirk *et al.*, 2007). It was noted that the 5'- and 3'-UTRs of mRNAs of eukaryotes and some RNA viruses like Hantavirus, Coxsackievirus B3 or Hepatitis C have different putative *cis*-regulatory elements or binding sites for proteins regarding the regulation of translation or mRNA stability (Luo *et al.*, 2003, Mir *et al.*, 2010, Verma *et al.*, 2010).

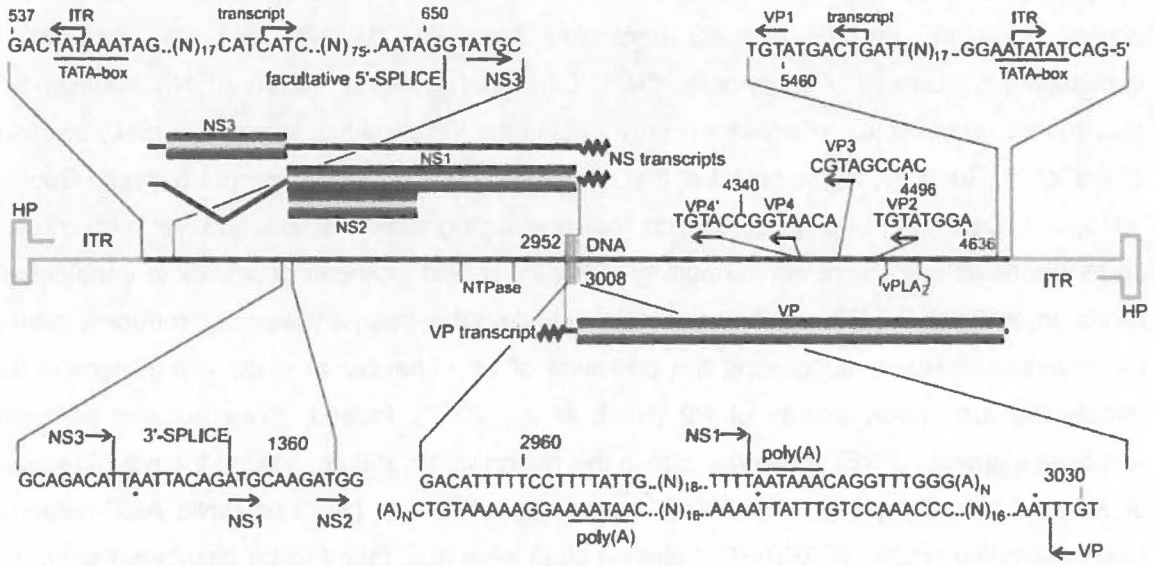


Figure 1-4. Details of MIDNV transcription map. Two putative promoter regions within the left and right ITRs control the expression of NS and VP transcripts. The 5'-splicing donor and 3'-splice acceptor sites are upstream NS3 and NS1, respectively. The two poly(A) signals are in the middle of the MIDNV genome. The initiation and termination of NS and VP transcripts are also indicated (Fediere *et al.*, 2004).

1.2.3.2. Transcription of *Pefudensovirus*

PfDNV transcription and VP protein expression was first studied by Yamagishi *et al.* (1999). This study reported the possibility of alternative splicing mechanism involved in VP expression by sequencing several cDNAs complementary to the VP transcripts. Sequencing results showed that alternatively spliced ORF1 and ORF2 were connected in frame to produce the largest VP transcript. Other VP transcripts with different termination sites were observed, presuming that this virus might use an alternative poly(A) signal to regulate VP transcripts (Yamagishi *et al.*, 1999). Apart from two sets of promoter P3 and P97 within the ITRs, another internal promoter region at map unit 18 was suggested responsible for the regulatory of NS1 and NS2 expression. Later, this promoter was proved a non-functional one (Yang *et al.*, 2008). Pefudensoviruses employ the same mode to express NS proteins as *Densovirus* members. Thus, NS3 was produced from unspliced transcript but NS1 and NS2 were generated after removal of the NS3-ORF by splicing and leaky scanning. It was noted that the last nt of Inr-box TCAGT is also the last nt of ITR sequence. Mutation of the Inr-box resulted in a complete drop of luciferase activity under P3 promoter control, confirming the importance of this motif in arthropods for RNA polymerase II during initiation steps of transcription in TATA-less promoters (Purnell *et al.*, 1994, Weis *et al.*, 1992). It was assumed that NS1 protein has a positive effect on its own promoter P3 in a concentration-dependent manner but not on the VP promoter. A $(CAC)_4$ repeat found in the PfDNV hairpin was probably related to the NS1 binding site for this *trans*-activation (Yang *et al.*, 2008, Yang *et al.*, 2006). This finding contrasts with previous results on *trans*-activation of the VP promoter by NS1 protein in vertebrate and invertebrate parvoviruses (Doerig *et al.*, 1990, Rhode *et al.*, 1987, Ward *et al.*, 2001b).

1.2.3.3. Transcription of *Iteravirus*

Iteravirus has three ORFs coding for NS1, NS2, and VP on the same strand. NS2-ORF is completely contained within NS1-ORF. Computer analysis revealed that iteravirus genomes may have different sets of promoter for each ORF containing either TATA-box (or TATA-like box) and Inr-box. The transcripts of VP and NS genes that would be generated from these promoters have not been identified. Expression of VP genes of CeDNV with different in-frame ATG but the same stop codon in baculovirus system yielded the identical structural proteins as native virus instead of four VPs in earlier reports (Fediere *et al.*, 2002, Nakagaki *et al.*, 1980). Based on the predicted splicing donor/acceptor sites, Kozak consensus sequences of five ATG

contexts and comparison of structural proteins expressed from baculovirus and from native virus, these authors suggested an alternative splicing for VP2 and VP4 and leaky scanning for VP3 and VP3'. To date, information regarding the transcription maps of viruses from this genus is lacking.

1.2.3.4. Transcription of *Brevidensovirus*

Brevidensovirus members have monosense DNA genome in which coding sequences are organized in three ORFs. Large, mid and right ORFs have potential coding capacities of about 700, 360, and 350 aa, respectively. The left ORF sequence contains conserved replication initiator motifs and NTP-binding and helicase domains similar to those in parvoviral NS1 proteins and thus most likely encodes for the Rep protein. Based on the analysis of conserved regions, there is a 33-35% aa identity between PstDNV and the AalDNV/AaeDNV while there is only a 17-19% aa identity between members of the *Brevidensovirus* and the lepidopteran densoviruses (JcDNV and BmDNV). Expression of the right ORF of PstDNV yielded virus-like particles (VLPs) implying that it codes for viral capsid protein (Kaufmann *et al.*, 2010) whereas the function of the protein coded by mid ORF is unknown.

Compared to other densoviruses, the transcription pattern of brevidensoviruses is not well characterized. AaeDNV and AalDNV do not contain any conserved donor or acceptor sequences, thus it was predicted that their mRNA transcripts are not spliced. Meanwhile, 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). However, these results appeared to be questionable for several reasons (see Publication 1, Section 1 in Results and Discussion). The activity of P2 and P61 was studied with and without transcriptional enhancer elements via the use of firefly luciferase reporter constructs in insect and fish cells. P2 promoters proved stronger in Sf9 cells and the two promoters were found to be constitutive promoters driving gene expression in both invertebrate and vertebrate hosts (Dhar *et al.*, 2007).

Among other brevidensoviruses, only the expression of AaeDNV has been somewhat documented. In early studies, β-galactosidase was fused in each of three ORFs and expression was obtained for NS2 and VP, but not for NS1 (Afanasiev *et al.*, 1994). It should be noted that, compared to AalDNV, AaeDNV was reported to have a 57 aa-extension at N-terminal of NS1 that could be non-existent and that the fusion protein was created after 47 aa in this extension (Afanasiev *et al.*, 1991, Boublik *et al.*, 1994). However, a fusion downstream of this N-terminal

extension yielded a good expression of NS1 (Kimmick *et al.*, 1998). Primer extension was used to detect the termination of VP transcripts at nt 2402 (corresponding to 2440 in AalDNV) and a trans-activation by NS1 protein could increase up to 10 times the VP promoter activity (Ward *et al.*, 2001b).

1.2.3.5. Transcription of *Cupidensovirus*

The only member of this genus is CpDNV, isolated from mosquito, but it differs from other mosquito densoviruses in both genome organization and transcription strategy. CpDNV has an ambisense genome and codes for three NS proteins on one strand and VP on the complementary strand as in the case of *Densovirus* members. However, each coding sequence of NS1 and NS2 is split into two smaller ORFs, thus both splicing and frameshifting were used in a way that put N-terminal sequence and C-terminal sequence of NS1 and NS2 in frame. There are three different promoters: P7 and P17 control the expression of NSs and P87 drives the expression of VPs. The two putative promoters P7 and P87 were detected upstream of the left and right ORFs, respectively, and are situated downstream the ITRs. This is different from other typical densoviruses that generally have ITR-overlapping promoters. These promoters were found to have both critical elements TATA-box and Inr-box (Baquerizo-Audiot *et al.*, 2009). Early reports showed that CpDNV has four structural proteins: two major proteins 64kDa-VP2 and 57kDa-VP3 and two minor ones 90kDa-VP1 and, possibly, 12kDa-VP4 (Jousset *et al.*, 2000). Since only one VP transcript encodes for four VPs with relative ratios suggested that leaky scanning by the 40S ribosomal subunit could occur during transcription. Like for *Densovirus* members that have 5nt-UTRs, the VP transcript of CpDNV has a very short 5'-UTR of only 3 nts, conducive to leaky scanning.

1.2.4. Role of non-structural proteins and palindromic telomeres in replication

1.2.4.1. Non-structural proteins

Densovirus genus members have three NS proteins, NS1, NS2, and NS3. NS1 is equivalent to the Rep protein (NS1) of vertebrate parvoviruses and localized in the nucleus. It is a multi-functional protein relevant to different steps within the viral cycle: nucleic acid replication, transcription regulation and virus assembly (Christensen *et al.*, 1995a, Christensen *et al.*, 1995b, Christensen *et al.*, 1993, Clemens *et al.*, 1990, Cotmore *et al.*, 1990). In the early steps of

replication, NS1 was found to covalently bind to an anchor or "tether" sequence. In mature virions, encapsidated ssDNA was found to be linked to NS1 protein via this oligonucleotide. This NS1-oligonucleotide complex was outside of the viral surface and cleaved during cell entry (Cotmore *et al.*, 1989b).

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, Fediere *et al.*, 2004, Im *et al.*, 1990, Tijssen *et al.*, 1995a, Tijssen *et al.*, 2003, Wilson *et al.*, 1991) (Figure 1-5). The RCR motifs found in NS1 proteins have two conserved clusters. The first, $UHUHU_3$ (U is a bulky hydrophobic residue), functions as a Mg^{2+}-binding site for nicking activity. The second, $UX2YUX_2K/R$, contains a tyrosine that acts as a nucleophile to break the phosphodiester bonds and the 3'-hydroxyl group (3'-OH) is liberated for strand displacement replication (Koonin *et al.*, 1993). The conserved ATP-binding and helicase motifs contain Walker-A $U_3XG/SP(X)_3GKT/SU$ and Walker-B $U_2X_2D/ED/E$ sites (Iyer *et al.*, 2001, Iyer *et al.*, 2004). Parvoviral helicases belong to the SF3 subfamily, which has 3'-5' translocation directionality and one ATP-binding domain functional in a multimeric form (Gorbalenya *et al.*, 1993, James *et al.*, 2003). Walker A is an ATP-binding loop or P-loop while Walker B is to coordinate Mg^{2+} and hydrolyze ATP for helicase activity. Densoviruses and iteraviruses have a GKN instead of consensus GKS/T residues in their Walker A-motif but their helicase was found to be active (Ding *et al.*, 2002). Recent studies also confirmed the role of NS1 in replication of JcDNV and PfDNV. It was reported that NS1 binds to the $(GAC)_4$ or $(CAC)_4$ motifs located inboard the terminal hairpin of JcDNV and PfDNV, respectively (Ding *et al.*, 2002, Yang *et al.*, 2006). In JcDNV, NS1 was demonstrated to perform nuclease activity at thymidines (5'-G*TA*TTG-3'). Meanwhile, in AAV2 and AAV4, the nicking was also found at thymidines but the cleavage sequences are divergent (Cotmore *et al.*, 1995b, Cotmore *et al.*, 2006). The nicking by NS1 is an important step to resolve replicative-intermediate (RF) DNA to monomeric daughter molecules. In *Densovirus* members, beside the enzymatic activity, NS1 protein is proposed to be a concentration-dependent regulator to the level of replicated DNA (Bergoin *et al.*, 2000). In the presence of low amounts of NS1 protein, there are limited numbers of multimeric intermediates of RF, leading to the accumulation of the double-stranded RFs, which can serve for transcription of more viral genes. In contrast, when NS1 protein is abundant, more nicks are created; subsequently more monomers of RF (mRFs) are produced that accumulate for viral encapsidation.

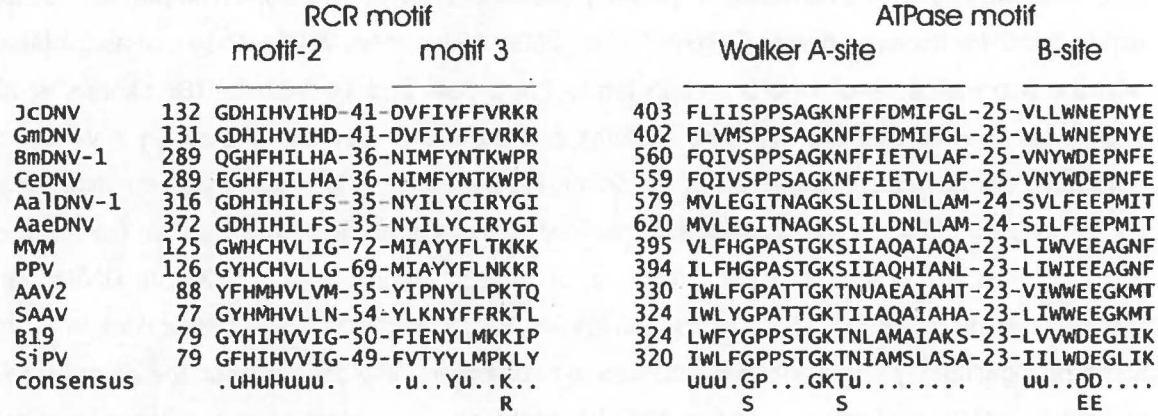


Figure 1-5. Conserved rolling circle replication (RCR) and Superfamily 3 (SF3) helicase motifs found in vertebrate and invertebrate parvovirus. Conserved residues in Densoviruses and Iteraviruses, the Walker A-site, contain GKN residues instead of GKS/T. *Densovirus* genus members usually have a single E residue in the conserved Walker B-site rather than double residues in other parvoviruses (Tijssen *et al.*, 2006).

Studies on vertebrate parvoviruses showed that, compared to NS1 protein, NS2 is more abundant and accumulates rapidly in the cytoplasm during early S-phase and has a shorter half-life, only about 1h (Cotmore *et al.*, 1990, Schoborg *et al.*, 1991). The NS2 in vertebrate parvovirus is believed to be important for efficient viral replication *in vivo* and *in vitro* (Cater *et al.*, 1992). Nonphosphorylated or both isoforms of NS2 were shown to interact with chromosomal region maintenance 1 (CRM1) (Bodendorf *et al.*, 1999), an important nuclear export factor for nuclear egress (Eichwald *et al.*, 2002, Miller *et al.*, 2002). They can also interact with the two members of 14-3-3 protein family (14-3-3eta and 14-3-3zeta) (Brockhaus *et al.*, 1996) that bind to different signaling proteins and therefore may be involved in a variety of regulatory processes in infected cells. Moreover, NS2 was found to interact with survival motor neuron protein (Smn) in the autonomous parvovirus-associated replication bodies (APARS) or Smn-APAR at late time of infection (Young *et al.*, 2002a, Young *et al.*, 2002b). In DNVs, very little information related to NS2 functions is known and, in fact, the NS2 polypeptides in DNVs share no significantly homologous sequences to each other. Only one study on the potential role of NS3 in JcDNV replication reported that viral replication was blocked in the absence of this protein (Abd-Alla *et al.*, 2004).

1.2.4.2. Rolling hairpin replication (RHR) model

Parvoviruses need to gain access to the host cell cytoplasm and then penetrate into the nucleus where replication of their genome occurs. Their ssDNA does not code for any cell cycle-modulation protein thus virus needs to wait until the cells enter S-phase (Cotmore *et al.*, 1987, Tattersall, 1972, Wolter *et al.*, 1980). Latent infection usually is not detected by the host innate immune systems. Many experimental results have rationalized the RHR model of parvovirus replication through duplex intermediates.

Parvoviral replication is a biphasic process. First, the free 3'-OH of the extreme nucleotide on the left hairpin is used as a primer by the DNA polymerase δ (Polδ) of the host replication machinery to synthesize a new complementary strand to the right hairpin of the original DNA (Baldauf *et al.*, 1997, Cossons *et al.*, 1996, Cotmore *et al.*, 1989a, Lochelt *et al.*, 1989, Tullis *et al.*, 1994) (Figure 1-6, step 1). This process is dependent on cycline A and associated kinase activity, which are normally necessary for cell cycle G1/S transition (Bashir *et al.*, 2000). The covalent duplex intermediate or covalently replicative form, cRF, then serves as a template for transcription to express NS proteins that are essential for viral replication of parvoviruses (Cotmore *et al.*, 1987, Deleu *et al.*, 1998, Naeger *et al.*, 1990).

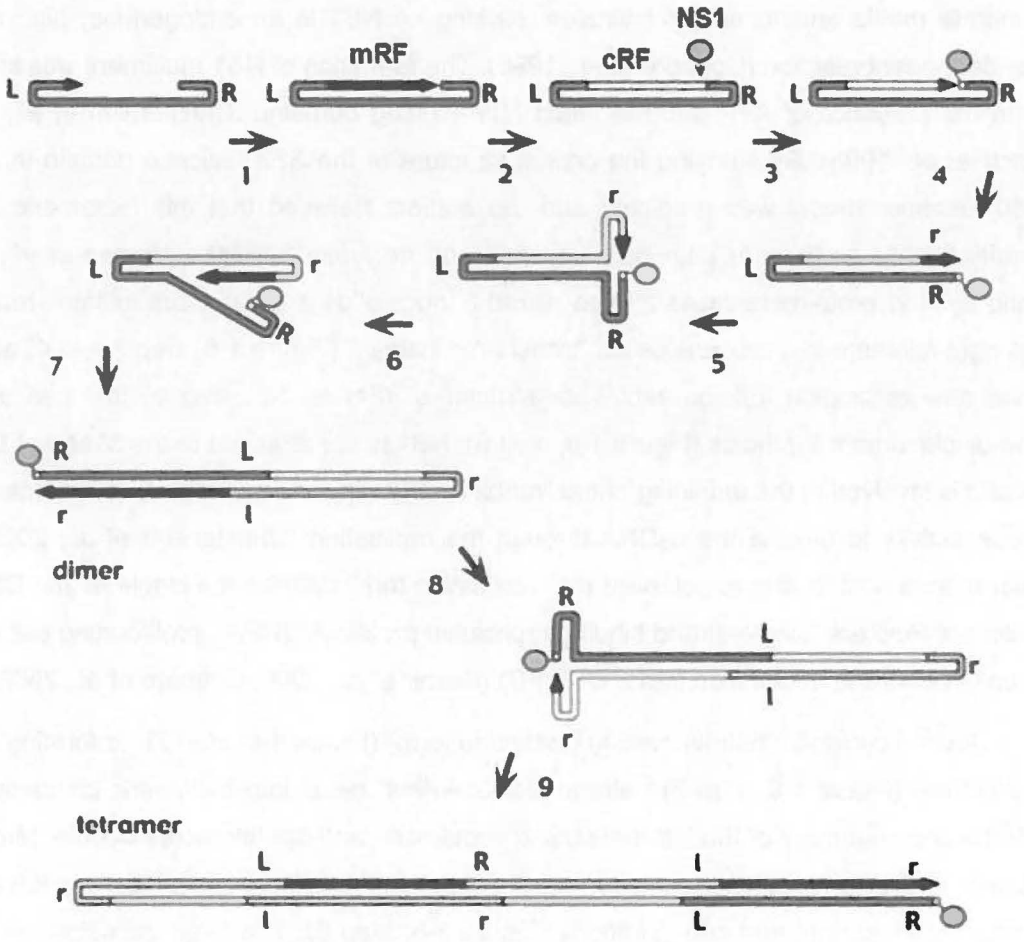


Figure 1-6. "Rolling hairpin replication" model of parvoviruses. L and R: left and right extremities in flip configuration; l and r: left and right extremities in flop configuration. The gray, white, and black strands represent for parental, progeny and newly synthesized genomes, respectively. The 3'-end is indicated by the arrowhead (Cotmore *et al.*, 2006)

Only NS1 in a multimeric form can recognize and bind to a specific site on the viral origins on the right telomere sequence (OriR), followed by site-specific endonuclease activity near the covalent link at the right hairpin (Figure 1-6, step 2) (Christensen *et al.*, 1995a, Cotmore *et al.*, 1995a, Cotmore *et al.*, 1995b, Nuesch *et al.*, 1995). Specific binding sequences can be degenerate motifs among different viruses. Nicking by NS1 is an endogenous, high mobility group-dependent reaction (Cotmore *et al.*, 1998). The formation of NS1 multimers was known to require the presence of ATP and the intact NTP-binding domains (Christensen *et al.*, 1995a, Nuesch *et al.*, 1993). By studying the crystal structure of the SF3 helicase domain in AAV, a Rep40 hexamer model was predicted and the authors believed that this hexameric form is essential for NS proteins to gain both helicase and nuclease activities (James *et al.*, 2003). Nicking by NS1 proteins releases a base-paired 3' nucleotide and therefore initiates replication of the right telomere in a process called "transferred hairpin" (Figure 1-6, step 3 and 4) and then creates new replication fork or "rabbit-ear structures" (Figure 1-6, step 5) that can serve in strand-displacement synthesis (Figure 1-6, step 6). NS1 is still attached to the 5' end of the viral DNA and is involved in the unfolding of the "rabbit ear" (Willwand *et al.*, 1997) or exhibits its 3'-5' helicase activity to unwind the dsDNA through the replication (Christensen *et al.*, 2002). The cellular factors contributing to establish the "replication fork" stabilize the single strand DNA and cofactors of Polδ are "single-strand binding replication protein A" (RPA), proliferating cell nuclear antigen (PCNA) and "replication factor C" (RFC) (Bashir *et al.*, 2000, Cotmore *et al.*, 2006).

Several cycles of "hairpin melting (extended form) (Figure 1-6, step 7) - reforming (rabbit-ear structure) (Figure 1-6, step 8) - strand displacement" result into multimeric concatemers in which the copy numbers of the internal coding sequences and the telomeres double. Many viral DNA units or monomer genomes are conjugated via a single palindromic junctions in left end:left end and right end:right end configurations (Figure 1-6, step 9). The large concatemers will be nicked by NS1 exonuclease to release the monomer length, duplex turn-around forms, or mRF (result of step 1). For parvoviruses that have ITRs, both palindromic sequences will undergo a similar process known as terminal resolution, in which right end termini are extended and cleaved by NS1 as discussed above. Since parvoviral ITRs are imperfect, the progeny viruses contain both "flip" and "flop" orientations, resulting in the equal amounts of minus and plus strands after encapsidation. Nonetheless, for those parvoviruses with unrelated terminal sequences or bubble(s) on their left hairpin, replication mechanisms for the right end is named terminal resolution but for the more complex left end, it is called "junction resolution" or "asymmetric mechanism". Like OriR, OriL on the left telomere can be recognized by NS1 protein via specific sites but indeed, the stability of NS1-DNA complex on the left palindrome sequence

requires the binding of the cellular parvovirus initiation factor (PIF) or glucocorticoid modulating element-binding protein (GEMB) (Christensen *et al.*, 1997, Cotmore *et al.*, 2006). It has been noted that in most vertebrate parvoviruses that have disparate telomeres, their left ends often contain PIF-binding sequence (Cotmore *et al.*, 2006). The asymmetric resolution generates positive strands that could be detected in replication pool but are not packaged. Meanwhile, terminal resolution for the right hairpin is more efficient and results in the preferential synthesis of minus strands for virus encapsidation by a "preferential strand displacement" or "kinetic hairpin transfer" model (Cotmore *et al.*, 2005, Cotmore *et al.*, 2006).

1.2.4.3. Densovirus telomeres are divergent in structure and sequence

Densovirus genomes share common features with autonomous vertebrate parvoviruses. Therefore, a similar replication model of vertebrate parvovirus can be proposed for DNV. Despite the diversity in the telomere structures of DNVs, these palindromic sequences at the termini can fold into Y-shaped, J-shaped, T-shaped or I-shaped hairpin (Figure 1-7) and serve as primers for viral replication at the 3'-extremity. However, the significant variety in their replication origins suggested particular distinguishing details for the different genera. Members of the *Densovirus* genus have long ITRs in which the Ori may be positioned at nt 95. These viruses share terminal structures that resemble those of AAV in flip and flop orientations and also contain the recognition site for terminal resolution in their palindromic sequence. On the other hand, iteraviruses seem to have both strands encapsidated with equimolar since they have similar ITRs but actually, "flip" configurations are found to be preferentially packaged in the virions. The ACCA motifs found on the ITRs of iteraviruses are similar to that of Minute Virus of Mice (MVM) where the NS1 binds to proceed to RHR. Moreover, the side arms of their hairpin are imperfectly base-paired, suggested a possibility of asymmetric strand displacement in the replication of telomeres. Brevidensoviruses, as MVM and many autonomous parvoviruses that package predominantly negative strands, have a unique palindromic sequence at their 5'- and 3'-ends. This dissimilarity in the telomere sequences suggests different resolutions for their termini replication and can explain the preferential minus strand encapsidation of brevidensoviruses (Tijssen *et al.*, 1995b).

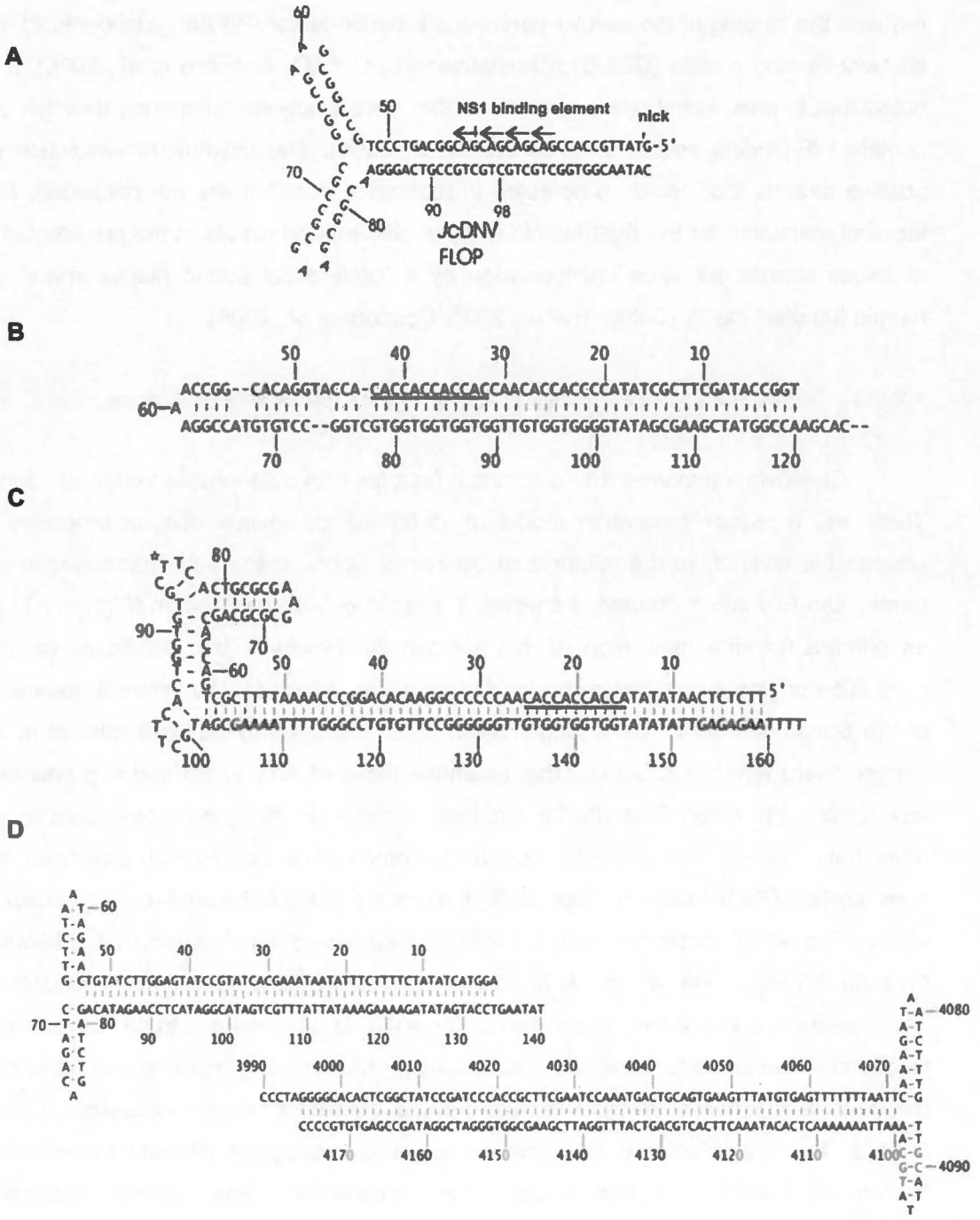


Figure 1-7. Typical hairpins of different genera of DNVs. Members of Densovirus genus with Y-shaped (A); Pefudensovirus with I-shaped (B); Iteravirus with J-shaped (C) hairpins. The T-shaped termini of Brevidensovirus are not identical and lack ITR (D). Potential NS1 binding sites are underlined (Tijssen *et al.*, 2006a).

1.2.5. Structural proteins and structure of densovirus

1.2.5.1. Structural proteins

Densovirus genomes are linear ssDNA, highly diversified from 4-6 kb in size and encapsidated in a symmetric icosahedral capsid. Like other parvoviruses, DNVs are non-enveloped and very small, from 18 to 25 nm, but extremely stable in a pH range from 3 to 9, 56°C for 60 minutes or in various non-polar organic solvents (Harris *et al.*, 1974). These characteristics of parvoviruses also explained why they could survive and still be infectious for a very long period in the environment. Filtrated or harvested virus stocks can be stored at cool temperature for years without losing infectivity, for examples; H-1, H-2, H-3, HB (different strains of H-1 rodent parvovirus) and feline panleucopenia virus (FPV) (Greene *et al.*, 1965, Poole, 1972, Toolan *et al.*, 1960). Some organic solvents such as chloroform-butanol or carbontetracloride can be used during the extraction step to improve virus yield and purity (Tijssen *et al.*, 1977, Tijssen *et al.*, 1976).

Ambisense DNVs have four forms of structural proteins, i.e. VP1, VP2, VP3, and VP4 with molecular masses ranging from about 45-90 kDa. For members of the *Densovirus* genus, these VPs are translated from a single ORF via leaky scanning with a ratio 14:26:21:39 (in % mass) in JcDNV (Fediere, 1983) and 1:9:9:41 (molecular ratio) in GmDNV (Tijssen *et al.*, 1977, Tijssen *et al.*, 1976). VP1 is a minor protein translated from the whole mRNA transcript. The smallest VP4 is the most abundant. Viral capsids can be assembled when only VP4 was expressed in baculovirus system (Croizier *et al.*, 2000, Fediere *et al.*, 2004, Tijssen *et al.*, 2003). PfDNV in *Pefudensovirus* genus has two ORFs encoding for VPs and it was reported to produce five VPs of 105, 82, 79, 56 and 52 kDa (Hu *et al.*, 1994, Yamagishi *et al.*, 1999). VP protein sequences between these two genera share less than 25% identity.

CpDNV is an ambisense densovirus but classified into another genus since this virus differs from others as previously described. Four VPs with molecular weights of 90, 64, 57, and 12 kDa were identified by SDS-PAGE analysis. VP2 and VP3 are major ones and the minor 12kDa-VP4 is not large enough to be a subunit for virus assembly (Jousset *et al.*, 2000).

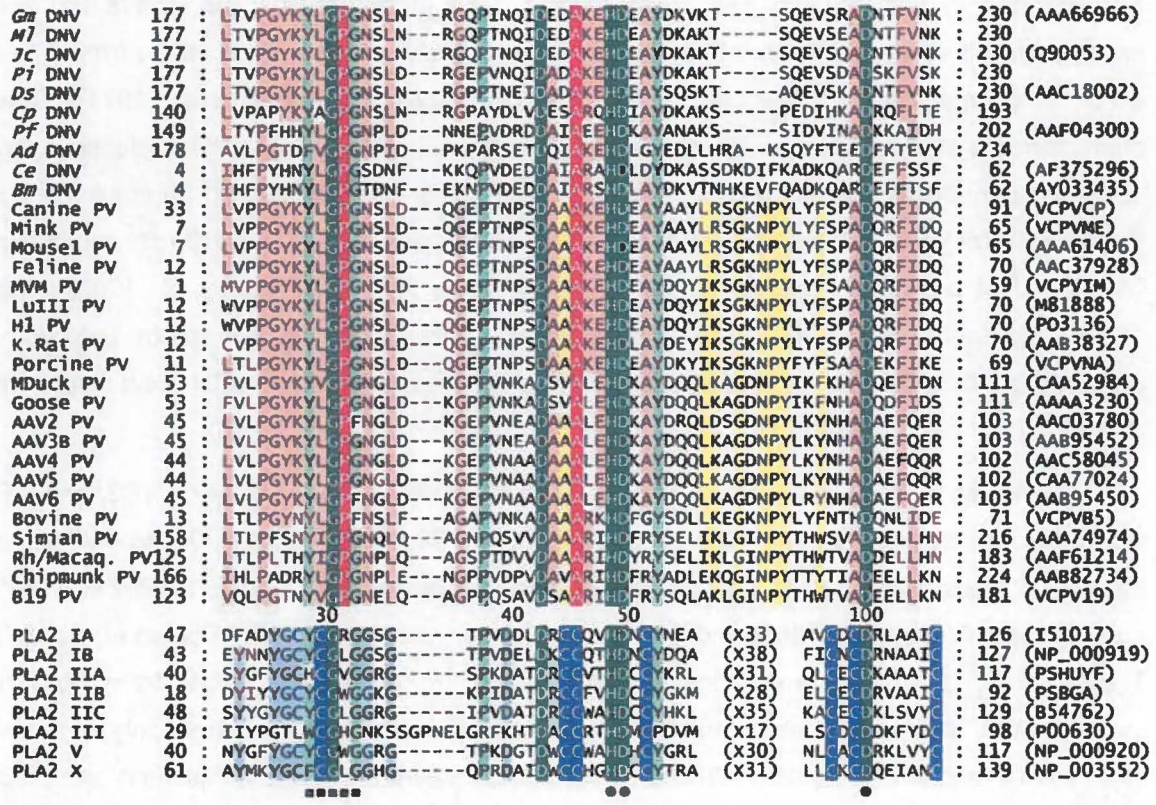


Figure 1-8. The conserved PLA2 motifs found in structural protein of different parvoviruses and in other sources (IA: Naja naja snake venom, IB: human pancreatic, IIA: human synovial fluid, IIB: gaboon viper snake venom, IIC: rat, III: bee venom, V: human PLA2, X: human. Dark green represents 100% identity; light green is at least 70% conserved among all PLA2s. Dark and light red is 100% identity and at least 70% conserved, respectively, among parvoviral PLA2s. Dark and light blue is 100% identity and at least 70% conserved, respectively, among secretory PLA2s. Yellow indicates conserved residues in vertebrate parvoviruses. Filled squares for conserved Ca^{2+} binding loop and filled circles for catalytic residues. GenBank numbers are listed on the right (Zadori *et al.*, 2001).

The smallest parvoviral capsid structure known is the shrimp brevidensovirus PstDNV, about 17 nm (Kaufmann *et al.*, 2010). Although previous reports demonstrated four VPs for PstDNV (Bonami *et al.*, 1990), only VP4 of 37.5 kDa is compatible with the estimated coding capacity. The recombinant VLPs of 37.5 kDa were successfully produced in baculovirus (Kaufmann *et al.*, 2010). AaeDNV and AalDNV are other brevidensoviruses but have two VPs with similar sizes about 40 kDa (Afanasiev *et al.*, 1991, Jousset *et al.*, 1993).

Nakagaki and Kawase (1980) reported four structural proteins of 77, 70, 57, and 50 kDa in BmDNV-1. Later, other reports showed four capsid proteins of 74.9, 64.3, 54.9, 51.6 kDa. The 57kDa-protein was shown to contain two isoforms of similar mass (Li *et al.*, 2001) and may derive from post-translational modifications. The VP3 was successfully expressed in baculovirus for near atomic structure study by X-ray crystallography (Kaufmann *et al.*, 2011). Like BmDNV, CeDNV also has five structural proteins with molecular weights of 76, 65, 56, 55, 48 kDa (Fediere *et al.*, 2002).

Different types of structural proteins of parvovirus share common C terminal sequences and differ only by the presence of their N-terminal extension (Tijssen *et al.*, 1981). In most parvoviruses, the largest VP1 is least present in the capsids. The upstream region of the VP1 protein or VP1up has been known to contain many important factors required for virus infectivity (Tullis *et al.*, 1993). These factors include conserved PLA2 domain and nuclear localization signals (NLS) (Tijssen *et al.*, 2006b, Vihinen-Ranta *et al.*, 2002, Zadori *et al.*, 2001). PLA2 is a Ca^{2+}-dependent enzyme and only catalyzes the hydrolysis reaction on aggregated substrates. Analysis of the PLA2 domain revealed the presence of two conserved motifs HDXXY and XYGXG for the catalytic site and the Ca^{2+} binding loop in both vertebrate and invertebrate parvoviruses (Figure 1-8) (Tijssen *et al.*, 2006b, Zadori *et al.*, 2001). These two regions can be located in VP1 upstream in iteraviruses and MVM or VP1 downstream in members of *Densovirus* and *Pefudensovirus* genera. Brevidensovirus, hepandensovirus and Aleutian mink disease virus (AMDV) are the only parvoviruses that do not have PLA2 motif. In vertebrate parvoviruses, during the entry step, virions need to undergo a rearrangement process in order to externalize the N-terminus of VP1 to the surface of the viral capsid (Canaan *et al.*, 2004, Cotmore *et al.*, 1999). The lipolytic activity of secretory PLA2 is then activated to breach the endosomal membrane (Farr *et al.*, 2005). It was reported that the PLA2 activity of PPV-VP1up or M33-VP1 up (truncated form of PPV-VP1) was 100-1000 times higher than those of AAV, B19 or GmDNV (Canaan *et al.*, 2004, Tijssen *et al.*, 2006b, Zadori *et al.*, 2001). The activity of PLA2 of Densoviruses is about 25-fold lower than that of Iteraviruses (Fediere *et al.*, 2004,

Fediere *et al.*, 2002, Li *et al.*, 2001, Tijssen *et al.*, 2003). Site-directed mutation of critical aa in PLA2 domains in PPV and deletion VP1up in JcDNV reduced both enzyme activity and virus infectivity (Abd-Alla, 2003, Zadori *et al.*, 2001) up to 100,000-fold. PLA2-mutated PPV was found to accumulate in the late endosomal/lyposomal compartments (Farr *et al.*, 2005, Zadori *et al.*, 2001).

Other conserved motifs among autonomous parvoviruses include GYKYL in the largest VP and a glycine/serine-rich region near the N-terminus of VP2 in parvovirus or VP4 in densovirus. The GYKYL motif is a 39-aa stretch with high identity between parvoviruses. The glycine-rich region is present in the hydrophobic helix in viral capsid structures (Tijssen *et al.*, 1995b, Tsao *et al.*, 1991, Xie *et al.*, 1996; Simpson *et al.*, 1998) (Figure 1-9). This region is positioned along the fivefold channel and flexible in order to pass through the fivefold channel during virus infection (Agbandje-McKenna *et al.*, 1998, Kontou *et al.*, 2005, Tsao *et al.*, 1991).

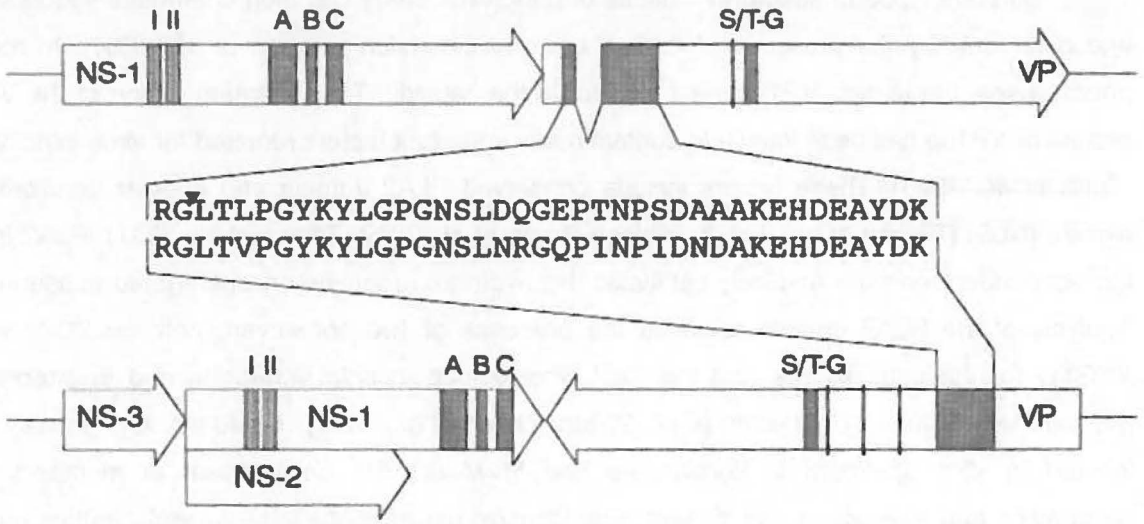


Figure 1-9. Comparison of conserved GYKYL and S/T, G-rich or glycine-rich motifs in structural protein of GmDNV (bottom) and PPV (top). GYKYL motif is disrupted after splicing in vertebrate parvoviruses but not in densoviruses. I and II: Rolling circle replication motifs. A, B and C: NS1 helicase superfamily III motifs (Tijssen *et al.*, 1995a).

1.2.5.2. Atomic capsid structure towards relating functions

Many near-atomic structures of full or empty capsid of parvoviruses have been resolved by X-ray crystallography since 1991 including members of *Parvovirus* (Agbandje-McKenna *et al.*, 1998, Tsao *et al.*, 1991), *Dependovirus* (Xie *et al.*, 2002), *Erythrovirus* (Kaufmann *et al.*, 2004), *Densovirus* (Simpson *et al.*, 1998), *Brevidensovirus* (Kaufmann *et al.*, 2010) and *Iteravirus* (Kaufmann *et al.*, 2011). By site-directed mutagenesis in the infectious clones, the available parvovirus structures allow us to link viral properties to structure and to understand capsid functions such as binding to host cell receptors, tissue tropism, antigenicity, pathogenicity, and evolutionary pathways. The icosahedral parvovirus T=1 capsid is composed of 60 single copies (capsomeres) of a common protein, usually the major protein (Figure 1-10A) (Caspar *et al.*, 1962). The major protein is VP2 in CPV, VP4 for GmDNV, and VP3 for BmDNV. N-terminal extensions are not part of the basic capsid structure, but have additional functions. These extensions are not ordered and are present only on a minority of the capsid proteins and thus cannot be observed by X-ray crystallography since their varied arrangement remains untraceable (Chapman *et al.*, 2006). 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 1-10B). This β-barrel structure is highly conserved in many viral families (Nandhagopal *et al.*, 2002, Rossmann *et al.*, 1989).

Eight β-strands are arranged into two four-stranded sheets, BIDG and CHEF (Figure 1-10D). The BIDG sheet is longer and closer to the inner surface than the CHEF. Meanwhile, two strands, βB and βD, are directed to the fivefold axis in contrast to βI and βG. In the vertebrate parvovirus CPV, an extra βA strand at the N-terminus folds back making discontinuous hydrogen bonds with βB in the same BIDG sheet (Chapman *et al.*, 2006, Xie *et al.*, 1996). However, in some densoviruses, GmDNV, PstDNV and BmDNV, this βA is a straightly extended form of βB (Figure 1-10D) and reaches to twofold axis to become anti-parallel with βB of the neighboring subunit. In GmDNV, hydrogen bonds were proposed for the interaction between βA and βB of other subunit and are thought to contribute to the capsid integrity (Simpson *et al.*, 1998). This exchange between two subunits creates a "swapping domain", a common feature found in insect viruses (Figure 1-11) (Kaufmann *et al.*, 2010, Kaufmann *et al.*, 2011, Simpson *et al.*, 1998).

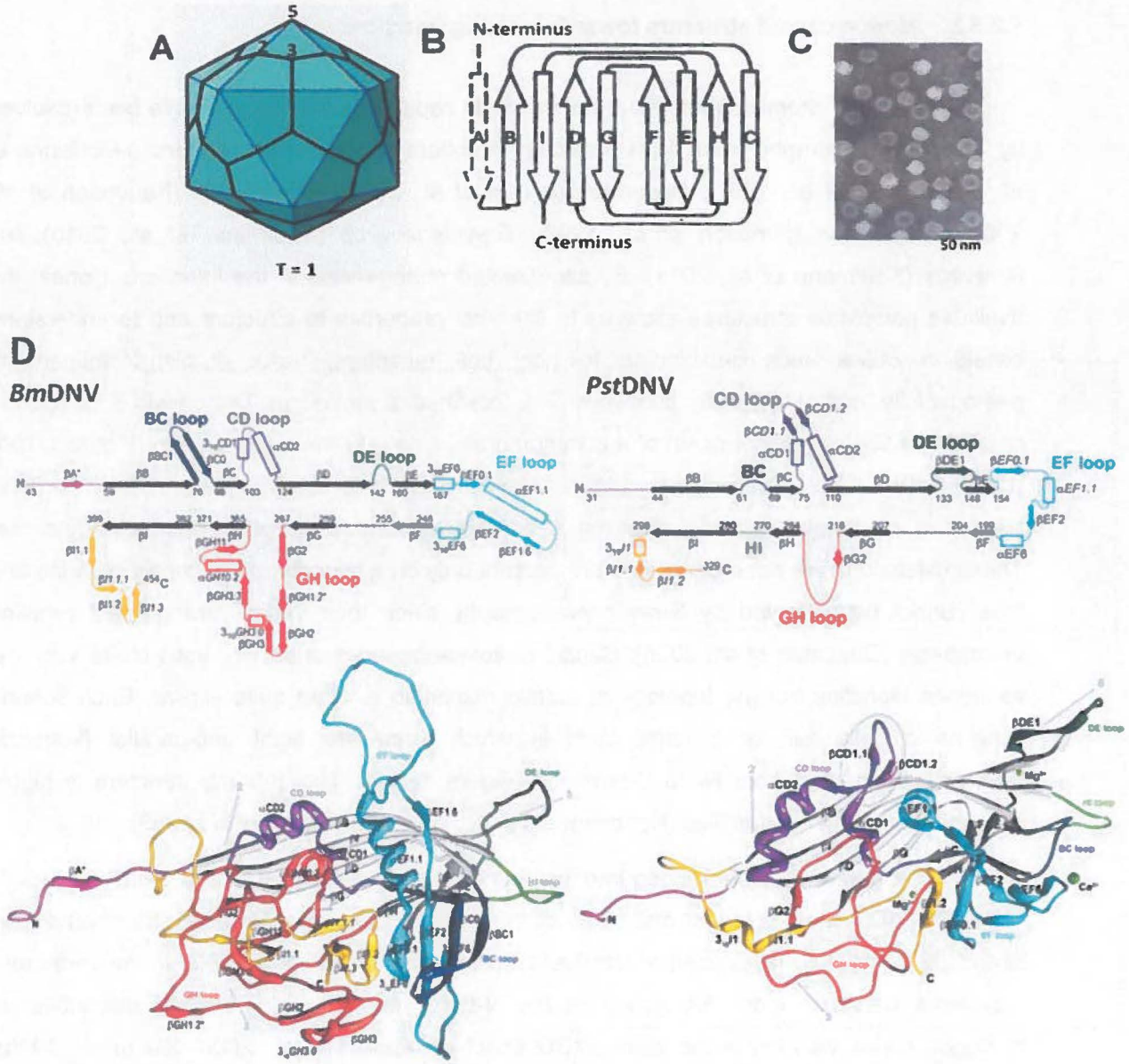


Figure 1-10. The structure of parvovirus capsid proteins. (A). Icosahedral capsid (T = 1) consiste of 60 identical subunits with 5-, 3- and 2- fold axes (ViralZone). (B). Topology arrangement of eight β-strands into a jellyroll β-barrels containing two β-sheet BIDG and FEHC in each protein subunit. βA exists in some but not all parvovirus structure (Chapman *et al.*, 2006). (C). Parvoviral virions visualized by transmission electron microscope (TEM) by negative staining method (Tijssen *et al.*, 2011). (D). Ribbon diagram of crystallized capsid subunit of BmDNV-1 (VP3, left) and PstDNV (VP4, right) with the ribbon diagrams showing the organization of β-strands and loops from N- to C-terminus in the polypeptide strands (top). The conserved β-barrel is in black (solid arrows) and the loops are in colors. Loops contain 70-80% of the polypeptide length of capsid proteins. They are divergent among parvoviruses and contribute to the parvoviral surface (Kaufmann *et al.*, 2010, Kaufmann *et al.*, 2011).

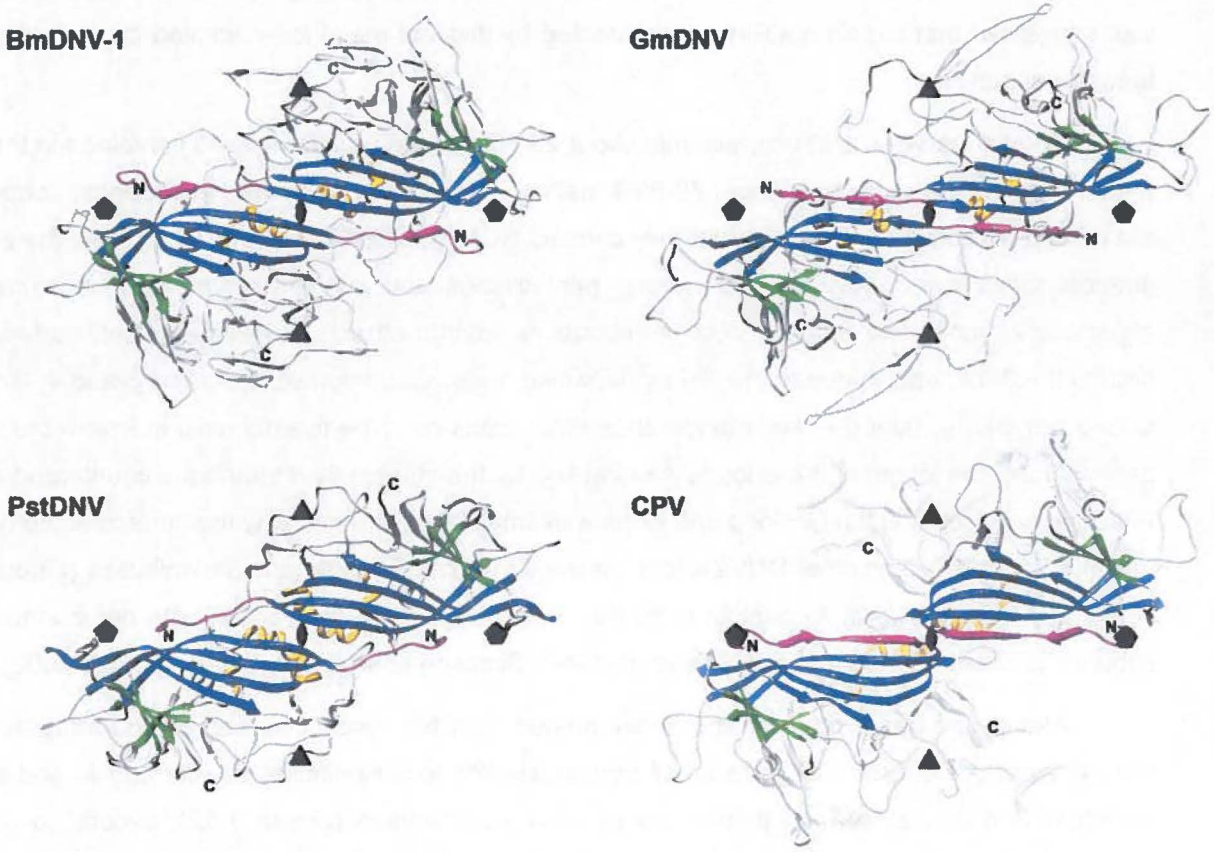


Figure 1-11. Illustration of "domain swapping" of N-terminus between two subunits in densovirus viewed at twofold axis. N-terminus is an extension of βA strand and interacts with the neighbouring βA strand via twofold axis in PstDNV, GmDNV, and BmDNV-1. In CPV and other vertebrate parvoviruses, the N-terminus folds back and becomes anti-parallel with βA strand within the same subunit. β-BIDG and β-CHEF sheets are in green and blue, respectively. Helical element of C-terminus is coloured yellow. N-terminal and βA is coloured magenta. The pentagon, triangle and ellipse represent for 5-, 3- and 2-fold axis (Kaufmann *et al.*, 2010).

Unlike GmDNV, in BmDNV, the βA is found to be anti-parallel with β-BIDG of neighboring threefold subunit; therefore, βA only makes two hydrogen bonds with the βB of twofold-related subunit. However, in PstDNV, there are no detectable hydrogen bonds and it was suggested that capsid stability was promoted by divalent metal ion-mediated cross-linking between subunits.

In most parvovirus structures, only about 20-30% of the protein mass is contained in the β-strands while the remaining; about 70-80% makes up the loops connecting the strands. Loops are named after the two β-strands that they connect BC, CD, DE, EF, FG, and GH. Unlike the β-strands, loops are known to differ among parvoviruses and account for surface structure, especially at the 3-fold apices. Most vertebrate parvovirus structures have a raised surface around the 3-fold apices due to the GH loops, which are divided into two sub-loops 3 and 4. GH loop is remarkably over ten times bigger than other loops near the fivefold axis. In invertebrate parvoviruses, the length of these loops is found shorter than that in their vertebrate counterparts. A lacking sub-loop 4 in the GH loop and loops with small length explain why the surface at the 3-fold axis in GmDNV and other DNVs is less elevated than that in vertebrate parvoviruses (Figure 1-12). GH loop is known to contribute to the capsid stability by interaction with neighboring subunits to create trimers (Llamas-Saiz *et al.*, 1997, Simpson *et al.*, 1998, Simpson *et al.*, 2002).

Among the densovirus structures determined, PstDNV reveals several unique features: first, all loops are shorter due to its small protein, leading to a significant smaller capsid and a smoother and thinner surface than those of other parvoviruses (Figure 1-12); second, a β-hairpin of unknown function containing βCD1.1 and βCD1.2 was inserted in conserved helical elements αCD1 and αCD2 (Figure 1-10D) (Kaufmann *et al.*, 2010). Moreover, there are some similar aspects found in PstDNV and B19 structure. The residues at the N-terminus are tending to the inner surface and the size of the 5-fold pore made by glycine-rich DE loop is narrow, about 5.1Å (Kaufmann *et al.*, 2004). This suggests that the fivefold pore can support the migration of DNA/proteins during genome packaging or early steps of infection as well as nuclear trafficking.

In general, the morphological structures of parvoviruses are similar: elevated 5-fold and regions around 3-fold axes, depressions at the 2-fold axes and around the concave regions of the 5-fold axes (Figure 1-12). The different observations gained from structure studies could reflect the evolutionary relationships among parvoviruses. PstDNV diverged from both insect densoviruses GmDNV and BmDNV.

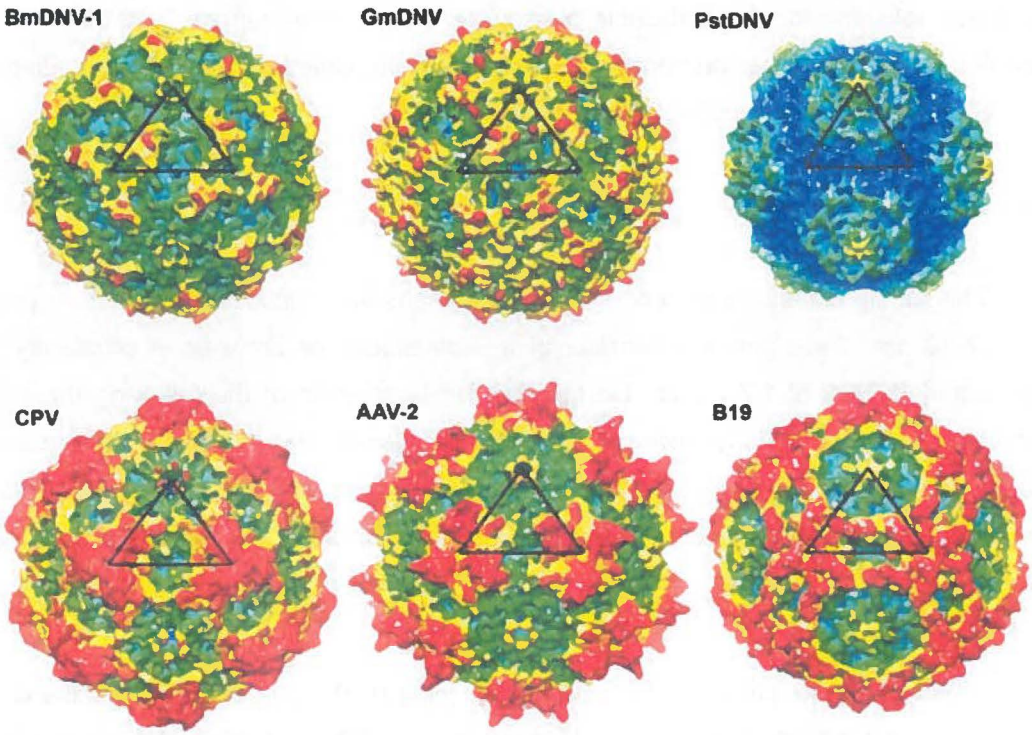


Figure 1-12. Comparison of surface topology between invertebrate and vertebrate parvoviruses. All surface structures were generated from the electron density maps at 8Å resolution. The colors are accounted for the distance from capsid center (blue, 100 Å; cyan, 107.5 Å; green, 115 Å; yellow, 122.5 Å; red, 130 Å). The triangle represents for one asymmetric capsid subunit (Kaufmann *et al.*, 2011).

1.3. Animal circular single-stranded DNA viruses (CSS-DNA viruses)

Based on their host range, CSS-DNA viruses are divided into at least four different families: *Geminiviridae*, *Nanoviridae*, *Circoviridae,* and *Anelloviridae*. Members of *Circoviridae* and *Anelloviridae* infect animals and humans while *Geminiviridae* and *Nanoviridae* infect plants. Unlike plant viruses, especially geminiviruses, which have been well documented, little information relevant to the molecular biology of animal and human CSS-DNA viruses is available. Those containing the conserved Rep protein are called circular Rep-encoding ssDNA (CRESS-DNA) viruses (CRESS-DNA viruses).

1.3.1. *Circoviridae* family

This family encompasses non-enveloped icosahedral animal viruses with a diameter of about 17-20 nm. Their genome consists of a monosense, or ambisense covalently circular single-stranded DNA of 1.7-2.3 kb. So far, this family consists of the following three genera: *Circovirus*, *Gyrovirus*, and *Cyclovirus*. Despite a high genetic diversity between these viruses, even in the same genus, the overall genomic architecture is well conserved. Members of the genus *Gyrovirus* differ from those of *Circovirus* and *Cyclovirus* by having a negative monosense genome and lacking the gene coding for rolling circle replication initiator protein (Rep) (Schat, 2009).

To date, only the chicken infectious anemia virus (CAV gyrovirus) and porcine circovirus (PCV) have been investigated in detail. Meanwhile, not all animal CSS-DNA viruses detected through metagenomics studies from animal tissue are correlated with specific diseases (Li *et al.*, 2011, Lorincz *et al.*, 2011, Lorincz *et al.*, 2012). Moreover, results from viral metagenomic also reveal a cross-species transmission of many CRESS-DNA viruses through the digestive tracts or faecal specimens whether an enteric virus replication does occur or not (Li *et al.*, 2010, Li *et al.*, 2011, Tan le *et al.*, 2013). Further studies need to be done to identify the real hosts and pathogenicity of these viruses to prevent potentially emerging epidemics in future.

1.3.1.1. *Gyrovirus* genus: non Rep-containing circular ssDNA virus

The first identified gyrovirus was the chicken anemia virus isolated from chicken presenting delayed growth, anemia, abnormal feathers, and leg paralysis as main symptoms after having been vaccinated against Marek's disease (herpesvirus) in Japan (Yuasa *et al.*,

1979). It was shown that the virus existed in US since early 1959 (Toro *et al.*, 2006). CAV is ubiquitous worldwide and all reported isolates have been known to be related to CAV diseases. To date, most CAV strains reported have been considered to belong to the same seroptype (Cardona *et al.*, 2000, Islam *et al.*, 2002, Nogueira *et al.*, 2007, Rosenberger *et al.*, 1989, K. A. Schat, 2009). This virus also caused immunodeficiency disorders resulting in gangrenous dermatitis and bone marrow aplasia in chicks (Goryo *et al.*, 1985, Yuasa *et al.*, 1979). However, infected adult chickens rarely show clinical symptoms but have a weakened immune system and become susceptible to other pathogens (Adair, 2000, Markowski-Grimsrud *et al.*, 2003). The immunopression caused by CAV can result in the decrease of vaccine efficacy. The complex interaction between CAV and many avian diseases caused by Marek's disease virus (MDV), infectious bursal disease virus (IBDV) or Reticuloendotheliosis virus (REV) could lead to higher mortality in infected chicks. Therefore, a combined vaccine has been suggested to be useful to protect against CAV. Although live vaccines have been developed and showed to be efficacious against CAV, it was later reported that the uses of these vaccines might be linked to the induced immunosuppression in chicks (Schat *et al.*, 2008).

The CAV genome has a negative sense of about 2.3 kb coding for the three proteins VP1, VP2 and VP3 of 51.6 kDa, 24.0 kDa, and 13.6 kDa, respectively (Figure 1-13) (Claessens *et al.*, 1991, Noteborn *et al.*, 1991, Noteborn *et al.*, 1995, Zhang *et al.*, 2012). Unlike other circular DNA viruses bearing Rep or Rep-like ORFs, CAV does not have any Rep protein coding sequence. However, as in geminiviruses and circoviruses, CAV also has a conserved GC-rich region including a stem-loop structure related to the initiation of DNA replication (Figure 1-13) (Rosario *et al.*, 2012b). This implies a similar RCR replication mode for CAV. Suprisingly, a 36-nt stretch in CAV from nt 2237 to 2272 (M55918) shares 80.6% sequence homology with nt 3816 to 3851 of Torque teno virus (TTV or Transfusion Transmitted Virus). Indeed, overall genome organization and positions of promoter, poly(A) signal, putative transcription factor binding sites such as the activating protein 2 (AP-2), the activating transcription factors/cAMP response element binding (ATF/CREB) and the nuclear factor kappa-light-chain-enhancer of activated B cells (NF-κB), are similarly found in CAV and TTV (Miyata *et al.*, 1999).

The precise transcription map and transcriptional strategy involved have not been resolved. However, it was showed that VP1, VP2, and VP3 might be expressed from a single, unspliced 2.1 kb mRNA by using different alternative start codons under the control of a single promoter and one poly(A) signal (Noteborn *et al.*, 1992, Phenix *et al.*, 1994). In addition to the major 2 kb-transcript, other spliced transcripts were detected during late transcription (Kamada

et al., 2006). Whether they are translated and which role they play in virus life cycle are still uncertain.

VP1 is the major capsid protein and is produced after VP2 and VP3 at 30h post infection (p.i.) (Douglas *et al.*, 1995). Evidence showed that VP1 and VP2 bind directly to each other and co-expression of both proteins is needed for a protective immune response (Koch *et al.*, 1995, Noteborn *et al.*, 1998). Since VP1 contains neutralizing epitopes, a suggested scaffold-like or chaperone activity was proposed for VP2. Indeed, VP2 has a signature motif $CXC(X)_3RK$ which is relevant to dual specificity protein phosphatase (DSP), including PTPase, and S/T PPase activities alike TTV (Peters *et al.*, 2002). Site-directed mutations of K102D and R101G in this motif could remarkably reduce viral replication and cytopathogenicity, respectively (Peters *et al.*, 2006). Other studies showed that VP2 may be an essential component of DNA pre-replication complex since it binds to mini-chromosome maintenance protein 3 (MCM3) and contains a NLS downstream of dual-specificity phosphatase (DPS) regions (Cheng *et al.*, 2012). Meanwhile, early studies showed that CAV-induced apoptosis in chicken thymocytes *in vivo* and *in vitro* is mainly due to the VP3 and therefore its name is apoptin (Jeurissen *et al.*, 1992; Noteborn *et al.*, 1994). Later, *in vitro* experiments showed that the G2/M arrest and apoptosis were solely induced in various human tumor cells in a p53-independent and Bcl-2-insensitive manner (Danen-Van Oorschot *et al.*, 1997, Teodoro *et al.*, 2004, Zhuang *et al.*, 1995). In normal cells, apoptin was found in the cytoplasm whereas when cells were transformed, this protein localized to the nucleus. VP3 contains a NLS (NLS1 and NLS2) at C-terminus, from nt 82-88 and nt 111-121 and a nuclear export signal (NES). The nuclear localization of apoptin was reported to depend on DNA damage response signaling via interaction with cellular anaphase promoting complex/cyclosome (APC/C)-associated factor (Kucharski *et al.*, 2011).

Recently, a new gyrovirus was discovered in diseased chickens from Brazil named Avian gyrovirus 2 (AGV2) (Rijsewijk *et al.*, 2011). Variants of this virus were also detected in diseased chickens in the Netherlands (dos Santos *et al.*, 2012). Its genome shares about 40% identity with that of the original CAV. Many new tentative human gyroviruses have also been discovered in blood samples from healthy donors and persons having received transplant organs or infected with HIV (Biagini *et al.*, 2013, Chu *et al.*, 2012, Maggi *et al.*, 2012, Sauvage *et al.*, 2011). Recently, a new GyV4 gyrovirus was detected in chicken meat and human stool in Hong Kong, suggesting a cross-transmission between species of this virus (Chu *et al.*, 2012).

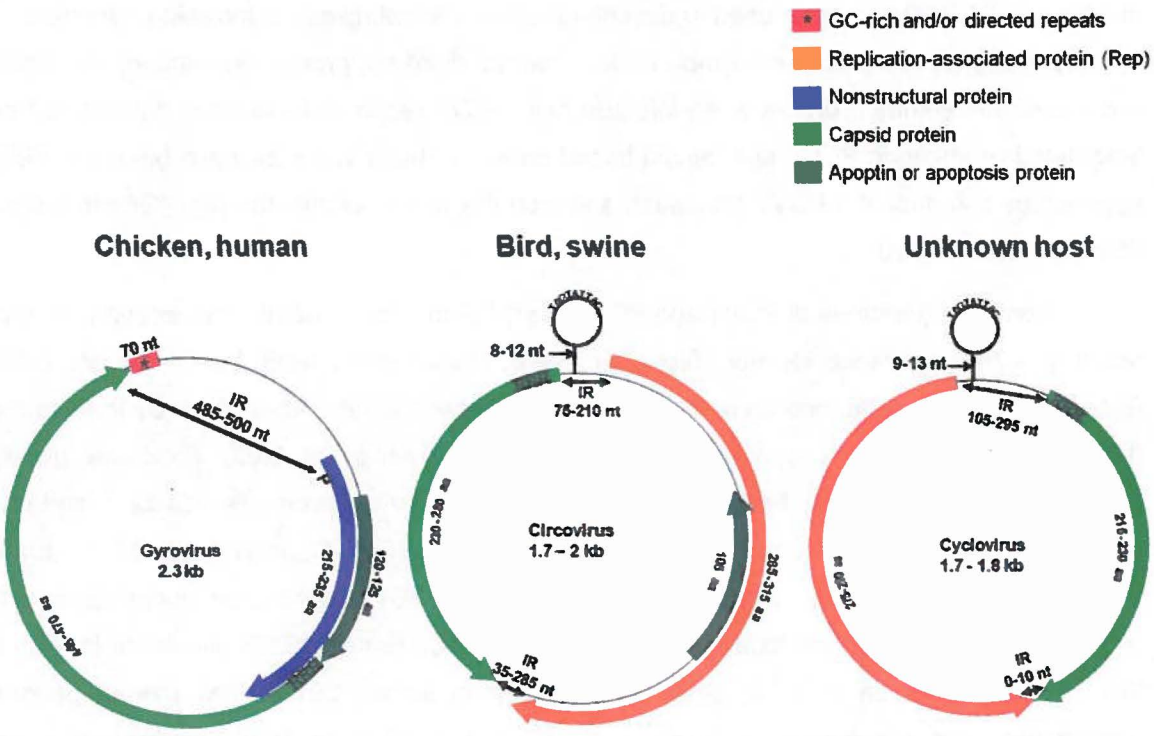


Figure 1-13. Genome organizations of Gyrovirus, Circovirus and Cyclovirus (modified from Rosario *et al.*, 2012b). IR: Intergenic regions. RRR: Basic aa-rich regions. P: Conserved phosphatase motifs. Arrows indicates the palindromic sequences in Circovirus and Cyclovirus.

1.3.1.2. *Circovirus* genus

A large number of circoviruses have been recently identified from a wide range of animal species including pigs, several domestic and wild birds (Delwart *et al.*, 2012, Halami *et al.*, 2008, Hattermann *et al.*, 2003, Mankertz *et al.*, 2000, Niagro *et al.*, 1998, Ritchie *et al.*, 1989, Todd *et al.*, 2001) and fishes (Lorincz *et al.*, 2011, Lorincz *et al.*, 2012). Phylogenetic analysis of the Rep sequences reveals that circoviruses might have evolved from geminiviruses, nanoviruses, and recombination with picornavirus in animal hosts (Gibbs *et al.*, 1999). PCV1 and PCV2 are the two circoviruses that have been well identified at the molecular level. PCV1 is a non-pathogenic virus isolated from a pig kidney cell culture PK15 (Allan *et al.*, 1995, Tischer *et al.*, 1986, Tischer *et al.*, 1974). PCV2 is associated with postweaning multisystemic wasting syndrome (PMWS) in young pigs in North America, Europe and Asia (Allan *et al.*, 1998, Choi *et al.*, 1999, Ellis *et al.*, 1998, Meehan *et al.*, 1998, Opriessnig *et al.*, 2007, Segales, 2012). PCV2 infects lymphoid tissue and causes lymphoid depletion and immunosuppression. Porcine circovirus associated

diseases or PCVAD has been used to describe multiple clinical diseases including PMWS, such as dyspnea, enlarged peripheral lymph nodes, icterus, diarrhea, growth retardation, etc. These syndromes are common on 6-8 week-infected pigs. PCV2 vaccines have been developed from inactivated or chimeric PCV2 and capsid-based subunit. These vaccines have been promising approaches to reduce the PCV2 circulation and mortality rate in vaccinated pigs (Gillespie *et al.*, 2009, O'Neill *et al.*, 2011).

Complete genomes of PCV1 and PCV2 are 1759 nts and 1768 nts, respectively, sharing about 68 - 76% sequence identity (Ellis *et al.*, 1998, Hamel *et al.*, 1998, Meehan *et al.*, 1998). Based on pairwise sequence comparison of capsid gene (ORF2) with a distance threshold of 0.035 or complete PCV2 genomes with a threshold distance of 0.02, worldwide genetic variations reported for PCV2 have led to the recognition of three genotypes PCV2a-c, and their genomes share 94.6-99% sequence identity (Cortey *et al.*, 2011, Gagnon *et al.*, 2007, Guo *et al.*, 2010, Zhai *et al.*, 2011). For PCV1, a second genotype PCV1-2a has been identified and its complete viral genome nucleotide shares about 86.4%, 88.7% and 86.5% sequence identity to that of PCV1, PCV2a and 2b, respectively (Gagnon *et al.*, 2010). The overall genome organization and replication mechanism of PCV2 are inherited from geminiviruses and nanoviruses since they share similar elements: a Rep-coding gene and the inverted repeats flanking the conserved octanucleotide (Oc8) motif. Oc8 motif encompasses a cleavage site $A_1X_2T_3A_4X_5T_6\downarrow A_7C_8$ (Oc8 or nonanucleotide motif TAXTAXT↓AC if one more T is included) for replication initiation (Cheung, 2012) (Figure 1-14A). It was found that the stem-loop configuration is essential for termination but not for cleavage during viral replication (Cheung, 2007).

PCV has an ambisense genome, which contains two ORFs coding regions for Rep (ORF1) on the plus strand and Cap proteins (ORF2) on the opposite strand (Figure 1-13). Two intergenic regions about 35-285 nts are located between ORF1 and ORF2. In PCV2, there are five potential ORFs coding for over 5 kDa proteins including the ORF3-encoded protein on the complementary strand (Meehan *et al.*, 1997). The ORF3-encoding protein was found to play an important role in the triggering of apoptosis via caspase-8 and caspase-3 pathway (Liu *et al.*, 2005). In PCV1, this protein is two times larger than that in PCV2 but has similar functions. This suggests that the apoptosis-induced activity itself does not determine the pathogenesis of PMWS caused by PCV2 in pigs (Chaiyakul *et al.*, 2010).

PCV1 and PCV2 utilize the spliced mechanism with comparable initiation and poly(A) signals to express their complex patterns of Rep genes. However, the expression level of their

transcripts and the splice junction used are different (Cheung, 2003a, Cheung, 2003c). The major 30 kDa-capsid protein (ORF2) and apoptin-related protein (ORF3) may have different promoters and were expressed from different, single transcripts (Nawagitgul *et al.*, 2000). Meanwhile, the Cap promoter was mapped within the exon of the ORF1 sequence (Mankertz *et al.*, 2004). The transcription profile of the Rep gene is very complex, comprising of Rep-related transcripts and NS-associated transcripts. Rep-related transcripts are unspliced mRNA which encodes for Rep protein whereas the Rep', Rep3a, Rep3b and Rep3c are products from mRNAs with different alternatively spliced junctions. Rep' is a C-terminal truncated form of Rep and both are controlled by the same promoter which is different from the one driving the expression of NS-associated transcripts (Mankertz *et al.*, 2004).

Rep and Rep' are the pivotal proteins for PCV replication *in vitro* (Cheung, 2006, Mankertz *et al.*, 2001). They have similar N-termini containing three motifs I (FTLNN), II (HLQGF) and III (YCSK) which are referred to typical initiator proteins for the RCR mechanism (Koonin *et al.*, 1993; Ilyina *et al.*, 1992). RCR-I motif is related to dsDNA binding site for Rep complex as in geminivirus and nanovirus (Gutierrez, 1999, Orozco *et al.*, 1998). Motif II is a divalent metal binding site involved in DNA nicking activity (Koonin *et al.*, 1993, Laufs *et al.*, 1995, Steinfeldt *et al.*, 2006, Vega-Rocha *et al.*, 2007). Motif III has a tyrosine residue containing an OH-group that makes a covalent link with the 5'-phosphate of the cleaved DNA strand (Laufs et al., 1995, Steinfeldt *et al.*, 2007, Vega-Rocha *et al.*, 2007). Since Rep' is a truncated form of Rep, another conserved motif, the GKS box or P-loop, is located only in the C-terminus of Rep but not in Rep'. Mutations of conserved aa on I, II and III motifs were shown to abolish the cleavage activity and therefore affect virus replication whilst a mutated P-loop reduced or completely abolished viral replication *in vitro* (Steinfeldt *et al.*, 2007). The expression of Rep and Rep' fused to GFP was detected in the nucleus but not in the nucleoli after transfection in PK15 cells. In PCV1-infected PK15 cells, both proteins could also be observed in the nucleus. This nuclear localization of Rep and Rep' was shown to depend on the three NLS, NLS1, NLS2, and NLS3 at their N-termini. In late infection, Cap proteins localized to the nucleoplasm and cytoplasm. However, in early infection, Cap proteins were found in the nucleoli but the function as well as nucleolar signal for this localization are still unknown (Finsterbusch *et al.*, 2005). The N-terminal 41 aa of Cap, in which two aa stretches 12RHRPRSH and 34HRYRWRRK41 were found to be strictly critical for the nuclear targeting of Cap proteins (Liu *et al.*, 2001).

The RCR model with the contribution of the Rep-Rep' hetero-complex has been accepted for PCV replication (Cheung, 2003b, Cheung, 2012). Like parvoviruses, PCV

replication in host cells is carried out during the S-phase and their ssDNA genome needs to be converted into dsDNA strand by the cellular factors. A similar mechanism prevails for replication of DNVs, PCV, geminivirus, and nanovirus. It was shown that Maize streak virus (MSV) and Banana bunchy top virus (BBTV) use the complementary oligos encapsidated in their virions as primers for second-strand DNA synthesis, while African Cassava mosaic virus uses a nascent complementary RNA molecule produced during the initiation step of viral infection as primers (Saunders *et al.*, 1992). However, so far, the primers for circoviruses have not been determined. The Rep-complex is known to bind to the Ori region at H1/H2 hexamers (Steinfeldt *et al.*, 2001) (Figure 1-15), thus allowing the unwinding of the dsDNA and exposure of the nonanucleotide motif for nicking activity by the Rep complex (Rep and Rep') (Steinfeldt *et al.*, 2006, Steinfeldt *et al.*, 2007). Nicking by the Rep complex generates a free 3'-OH serving as primer for RCR, while the Rep complex remains covalently bound to the parental genome. For replication of the palindromic sequences, a "melting pot" model was proposed rather than a "cruciform" one as for geminiviruses (Cheung, 2004a, Cheung, 2004b). In this model, binding of the Rep complex destabilizes the dsDNA and induces a sphere of four-stranded structures in which the hydrogen bonds between the palindromic sequences are in a "weakened" state. Therefore, both the minus and plus strands in inverted repeats are close enough to each other and can serve as template in template switching of the leading strand (Figure 1-15). Template switching of leading strands can occur in the initiation and termination steps of replication processes. It means that when the replication fork reaches to the inverted repeat sequences, two templates can be served for the synthesis of the right end and left end. Once a replication cycle is completed, the Rep complex may cleave at Oc8 between the newly synthesized strand and parental strand. The parental strand is then released and the Rep complex displaces to attach to the 5'-progeny strand, joins the 3'OH- and 5'-phosphate of this strand so that the nonanucleotide motif is reconstituted. This event results in a dsRF intermediate containing the newly synthesized strand and a single parental strand. The newly synthesized ssDNAs can be nicked again for next replication cycles to produce more DNA for encapsidation (Cheung, 2004a, Cheung, 2006, Cheung, 2012).

However, many aspects cannot be explained by this replication model. First, the primer for synthesis of double strand DNA at the beginning of replication is unknown; second, the role of Rep' in PCV replication in mammalian cells since in *E. coli*, the replication of PCV-based plasmid does require only Rep (Cheung, 2006); and finally, the "melting pot" mechanism cannot explain the observed reversion phenomenon of mutated Oc8 sequence (O1, O3 and O4) into wild-type sequence (Cheung, 2004b).

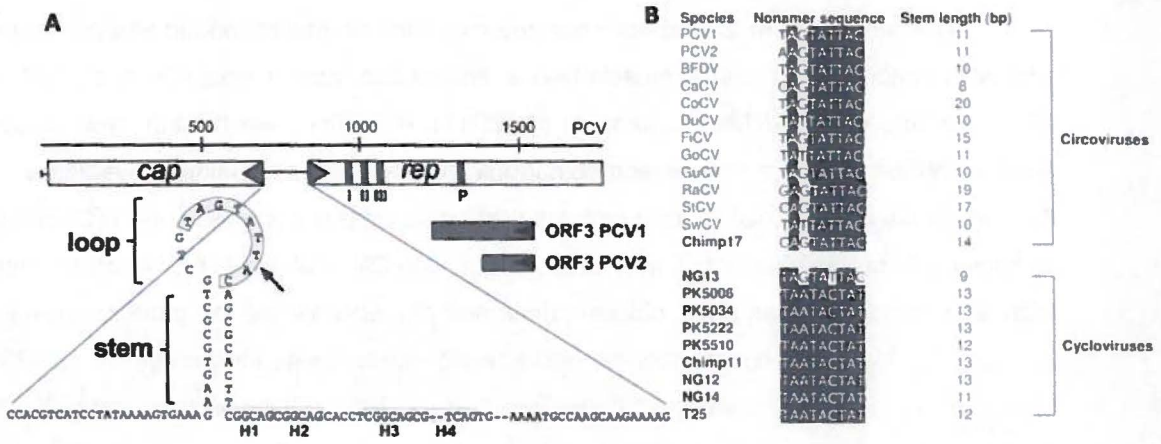


Figure 1-14. (A). Linear map of PCV with rep and cap coding sequences with conserved motifs I, II and III of RCR model. P is conserved P-loop for dNTP-binding. ORF3 of PCV1 has double size compared to that of PCV2. Stem-loop structure including conserved nonanucleotide motif is flanked by inverted repeats in PCV1 genome. Nick site is indicated by arrow. H1, H2, H3, and H4 are tandem repeats for binding of Rep-complex (Mankertz *et al.*, 2004); (B). Conserved nonanucleotide motifs in some circoviruses and cycloviruses (Li *et al.*, 2010).

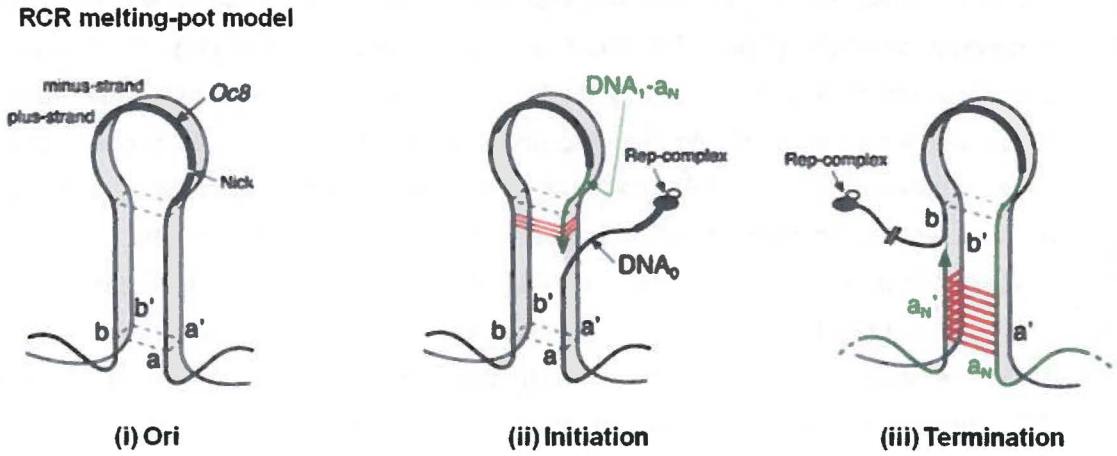


Figure 1-15. RCR "melting-pot" model for PCV replication. (i). The RCR "melting-pot" model with four-stranded structure of plus (a and b) and minus strands (a' and b'). (ii). In the initiation step, new synthesized DNA (DNA_1 in green) can use both a' and b as templates, allowing template switching to synthesize a_N strand. (III). In the termination step, a and b' can be used as templates for replication of a_N'strand. DNAo is parental DNA strand (in black). a_N' is complementary of a_N and b' showed by red lines (Cheung, 2012).

1.3.1.3. *Cyclovirus* genus

Most viruses from this genus were detected through metagenomic analysis from different dragonfly species, farm animal muscle tissue, animal and human stool (Ge *et al.*, 2011, Li *et al.*, 2010, Rosario *et al.*, 2012a, Rosario *et al.*, 2011). Recently, new human cycloviruses (CyVs) were identified in serum and cerebrospinal fluid (CSF) from unexplained paraplegia patients in Malawi (Smits *et al.*, 2013) or from patients with acute central nervous system (CNS) infection of unknown etiology in Vietnam (Tan le *et al.*, 2013). The CNS-related CyVs reported from Vietnam was also found in human stool, chicken, duck and pig feces where the patients were living (Tan le *et al.*, 2013). Nucleotide sequence similarities between these strains are about 97% so they were considered as the same genotype (Tan le *et al.*, 2013). Epidemiology and pathogenicity of human CyVs need to be investigated more deeply in order to link them to potentially emerging epidemic diseases.

Except for their genome sequence, information regarding the molecular biology of these viruses is still scarce. CyV genomes are closely related to those from members of the *Circovirus* genus. However, since they were grouped in a same cluster on Rep-based phylogenetic tree, a new genus *Cyclovirus* was proposed for them (Li *et al.*, 2010, Rosario *et al.*, 2012a). Their genome is small, about 1723-1867 nts, with two main ORFs encoding for Rep and Cap proteins in opposite directions (Figure 1-13). Unlike members of the *Circovirus* genus, in CyVs, the two ORFs are very close at the 3' end (0-10 nts), or can overlap in Dragonflies cyclovirus-5 (DfCyV-5), but are distant at the 5' end (105-295 nts) (Rosario *et al.*, 2012b). Sequence identity of Rep proteins between different CyV species is over 40%, whereas Cap proteins share very low or even no similarity to each other (Tan le *et al.*, 2013). The Rep gene contains also RCR and SF3 helicase motifs as found in circoviruses and other CRESS-DNA plant viruses. CyV genomes contain a stem-loop structure with a conserved nonamer sequence that is flanked by short inverted repeats of 9-13 nts (Figure 1-14B). These characteristics suggest a circovirus-similar RCR mechanism for CyV replication. In all circoviruses, the Rep gene has a small intron, implying a similar splicing mechanism for the expression of Rep and Rep' (Li *et al.*, 2010, Tan le *et al.*, 2013).

1.3.2. Unclassified novel viruses and new proposed taxonomy

The number of new Rep-bearing CSS-DNA viruses has exploded during the last decade in various environmental sources and extensive host range from human (anellovirus) (Biagini,

2004, Handa *et al.*, 2000), primate (Blinkova *et al.*, 2010, Li *et al.*, 2010), mammals (Allan *et al.*, 1998, Allan *et al.*, 1995, Kapoor *et al.*, 2012), birds (Halami *et al.*, 2008, Hattermann *et al.*, 2003, Niagro *et al.*, 1998, Ritchie *et al.*, 1989), insects (Dayaram *et al.*, 2013, Padilla-Rodriguez *et al.*, 2013), fishes (Lorincz *et al.*, 2011), planktons (Dunlap *et al.*, 2013, Yoon *et al.*, 2011), plants (geminivirus and nanovirus) (Abraham *et al.*, 2010, Andersen *et al.*, 1988, Bell *et al.*, 2002, Krupovic *et al.*, 2009, Roberts *et al.*, 1988, Stanley *et al.*, 1986, Timchenko *et al.*, 2006, Vega-Arreguin *et al.*, 2007) and phytoplasmas (Oshima *et al.*, 2001). Many of them have distinct genome architecture and bear an unknown Cap gene that is completely different from other existent CSS-DNA viruses. For example, Shan *et al.* in 2011 reported four Porcine-circo-like viruses i.e. 21, 22, 41, 51 isolated from pig fecal specimens. These viruses have rather big genomes of 3912, 3923, 2904 and 2823 nts containing at least three ORFs. Their Rep-related sequence is closest to the Rep gene that was integrated in the genomes of *Entamoeba histolytica* and *Entamoeba dispar* (about 33 to 34%), suggesting a new family for these viruses. Indeed, the position of a stem-loop structure can be different; downstream or upstream of the Rep gene (Shan *et al.*, 2011). Another study reported seventeen novel CSS-DNA viruses (RodSCVs) found in the fecal virome of rodents with genomes ranging from 1,124 nts to 3,781 nts. These viruses also have different kinds of genomic organization (Phan *et al.*, 2011). Except for the Rep-related ORF, the other ORFs in these genomes did not show sequence similarity to any available proteins on Genbank. Various CSS-DNA viruses with genomic diversity were also detected in dragonflies. Analysis of these viral genomes by Pfam revealed that one of these genomes contains an ORF coding for a Tombusvirus-like capsid protein, a genus of single stranded RNA viruses infecting plants (Rosario *et al.*, 2012a). Many novel uncharacterized anelloviruses and circoviruses recently detected from mosquitoes have mixed features of geminivirus and germinivirus-related DNA mycovirus. Some of these viruses have geminivirus-related Cap but circovirus-related Rep genes (Ng *et al.*, 2011).

These novel CSS-DNA viruses discovered through metagenomic reveal high diversity in genomic sequence and organization. They are not clustered into any existing groups by phylogenetic analysis and therefore, remain outside the recent classification system for CSS-DNA viruses and cannot be integrated in the *Circoviridae* family.

It should be stated that metagenomic analysis by deep sequencing might lead to artifacts (Naccache *et al.*, 2014, Zhi *et al.*, 2014) and that some of the studied animals (dragonflies, bats) are notorious collectors of non-host specific viruses.

Objectives

- Understanding the transcription pattern of two putative brevidensoviruses: PstDNV and AalDNV.

- Study of the expression strategy of AdDNV and geographycal/phylogenetic analysis of different AdDNV isolates.

- Cloning and sequencing of densoviral termini.

- Viral diagnosis to determine and characterize novel viruses in crickets and shrimps.

2. Materials and Methods

2.1. Cells and tissue culture

Cell type used in this study:

SF9: *Spodoptera frugiperda* ovarian cells (Lepidopteran cells)

C6/36: *Aedes albopictus* cells (Dipteran cells)

LD652: *Lymantria dispar* cells (Lepidopteran cells)

HeLa: human cervical cancer

293T: human embryonic kidney cells

Cell monolayers were generally grown in $25cm^2$ or $75cm^2$ tissue culture flasks in appropriate media (DMEM for HeLa and 293T, RPMI for C6/36 and Sf900 for Sf9 and LD652) supplemented with 10% (v/v) heat-inactivated FBS (foetal bovine serum). Human cell lines were trypsinised and regularly passed every 4-7 days and incubated at 37^0C in 5% CO_2. For insect cells, cells were detached and passed every 4-5 days by pipetting, diluted as appropriate and incubated at 28^0C.

2.2. Cloning of viruses and promoters

2.2.1. Virus clones

pCR2.1-PstDNV was obtained by transferring a sequence of 3873 nts genomic clone of PstDNV (Shike et al, 2000), that lacked the terminal hairpins, into the pCR2.1 vector (Invitrogen) at EcoRI sites. This PstDNV insert was also transferred into a baculovirus, via the pFasbac1 baculovirus expression vector (Invitrogen), from which the polyhedrin promoter was removed (Pst-baculovirus). For this purpose, a BamHI site was introduced into pFastBac1 at position 3870. Digestion of this vector with BamHI, at positions 3870 and 4032, removed the polyhedrin promoter (nt 3904-4032). The PstDNV genome was then subcloned into mutant pFastBacI between BamHI and XhoI sites.

pCR2.1-AalDNV was similarly obtained from a plasmid containing the complete genome of AalDNV of 4176 nt (Boublik et al, 1994) by transferring the insert into the pCR2.1 vector at EcoRI sites.

pSK(-)-CaAdVVV, pSK-JpAdVVV and pSK-USGaVVV: complete genome of AdVVVs isolated from *A.domestitus* in Canada, Japan and from *G.assimilis* in United states were cloned into the pSK(-) vector as described in Publications 7 and 8, Section 4, Results.

pSK(+)-PmCV-1 and pSK-PmaCV2: complete sequence of PmCV-1 and PmaCV-2 were cloned into the pSK(+) vector as described in Publication 9, Section 4, Results.

pKS(+)-AdMADV1 and pSK-AdMADV2: two clones containing each EcoRV-digested fragment of complete AdMADV genome were inserted into the pSK(+) vector as described in Publication 10, Section 4, Results.

pSK(+)-AdSDNV: complete NS coding sequence flanking by two hairpin structures of AdDNV2 was inserted into the pSK(+) vector as described in Publication 11, Section 4, Results.

2.2.2. Viral promoter constructs

Predicted promoter regions of PstDNV were amplified and cloned into the EcoRV site of the pGL4.20 basic vector (See Table 3-1, Publication 1, Section 1, Results for primer sequences). The AalDNV promoters P7/7.4 for the NS proteins were found to overlap each other (diagram 3-8B and 3-8C in Publication 2, Section 1, Results). Several constructs were made to dissect the critical elements for the expression of either NS1 or NS2 (see primer sequences on Table 2-1 below). In summary, fragment I (KpnI-SacI amplicon) and II (SacI-PstI amplicon) were generated by PCR and cloned into the pGL4 basic vector. Fragment III with the overhanging PstI and HindII ends was generated by primer hybridization (for NS1 promoter constructs) or by PCR (for NS2 promoter constructs) and ready for ligation into the same vector pGL4 basic that already contained fragment I and II (ATG of NS1 was in-frame with ATG of luciferase).

The same approach was used to introduce a stop codon for NS1/NS2 or to knock out the start codons of NS1 and NS2 but with primers, which contained desired mutations (see primer sequences on Table 2-1 below).

To knockout the start codon of the luciferase gene, forward primers with desired mutation was used in PCR to amplify the luc gene (including the SV40 poly(A) signal in the vector) and then ligated to all the previously made constructs (into HindIII and BamHI).

The putative TATA-box of NS2 (TATA2) was knocked out by replacing a XhoI sequence in designed primers. PCR was carried out on the F2 construct and products were cloned into the KpnI and XhoI sites of pGL4. The putative Inr2 of the NS2 transcript was knocked out by PCR with appropriately mutated primers using F2 and F5- as templates and cloning into XhoI and ApaI sites (downstream of the ATG of luc gene).

Table 2-1. Primers used for different constructs of NS1 and NS2 promoters of AalDNV

Primer names	Sequences	Purpose
A82F-Kpn	GCGGTACCGTATCTTGGAGTATCC	Segment I and I-III
A289R-SacI	AGCGAGCTCATGTGGTGGTGGATTATT	Frament I
A294F-SacI	AGCGAGCTCCAGAGTATATAAGAGGT	Segment II
A360R-PstI	GAATGTTCACTGCAGACTGATTCCAT	Segment II
A341F-PstI	CAGTCTGCAGTGAACATTCG	Segment III for NS2
A417R-HindIII	CAAGCTTGCGTCATGCTCCCACC	Segment III for NS2
A360R-PstI-ΔATG[1]	GAATGTTCACTGCAGACTGATTCCAAACTACTGACTC	Segment II - knockout ATG of NS1
A416R-HindIII-stopNS2	CCAAGCTTACGTCATGCTCCCAAC	Segment II - introduced TAA for NS2
A416R-HindIII-ΔATG[2]	CCAAGCTTGCGTCTTGCTCTTGCTCCCAAC	Segment III-knockout ATG of NS2
A342FPstI	GTGAACATTCGCGTGGGCA	Segment III for NS1
A342FPstI-stopNS1	GTGAACATTCGCGTGGGCTA	Segment III - introduced stop for NS1
A362R-HindIII	AGCTTGCCCACGCGAATGTTCACTGC	Segment III for NS1 (348-362)
A362R-HindIII-stopNS1	AGCTTACCCACGCGAATGTTCACTGC	Segment III - introduced stop for NS1
DF-HindIII-ΔATG[luc]	CTTAAGCTTGGCAATCCGGTACTGTTGGTAAAGCCACCTTGG	Segment IV - knockout ATG of luciferase gene
DR-ΔATG[luc]	AAGAGCGCCCAATACGCAAACGGAT	Segment IV - knockout ATG of luciferase gene
Akpn77F	GCG GTA CCG TAT CTT GGA GTA TCC	Segment II, knock-out of TATA2
Axho320R	CTT CCT CGA GCC CTC TTA TAT ACT CTG G	Segment II, knock-out of TATA2
Axho307F	AGG GCT CGA GGA AGG GAG AGT CAG TAG TAT GGA ATG CCG CTG CAG TGA ACA TTC	Segment II, knock-out of Inr2
Aapa496R	TAG AAT GGC GCT GGG CCC TTC TTA ATG	Segment II, knock-out of Inr2

2.3. Cytochrome b gene amplification to determine cricket species

Total DNA was extracted from the whole cricket using the standard proteinase K/phenol DNA isolation method. Five µl of extracted DNA was directly used as template for PCR in a 50 µl reaction which contains: 1 U Taq DNA polymerase, 1X PCR buffer, 0.25 mM dNTPs, 10 pmol of each primer. A 550 bp-fragment of cytochrome b gene was amplified with primers 2F (GTAATAGCAACAGCWTTTATAG) and 2R (CCWARTTTATTAGGAATAGATCG) (Huang*et al.*, 2000). PCR was performed by denaturing step at 95^0C for 5 min, 35 cycles of 95^0C (60 s), 50^0C (40 s) and 72^0C (1 min), and final extension at 72^0C for 5 min. PCR products were visualized on 1.5% agarose gels. Fragments were isolated from agarose gel for further purification. DNA sequencing was performed by Sanger's method using the same primers used for PCR.

2.4. Cell transfection for RNA isolation and luciferase assays

Before transfection, cells were grown for 24 hrs in 6- or 24-well plates for luciferase assays in appropriate medium without antibiotic. The DOTAP kit (Roche) was used for transfection of LD652 and Sf9 cells by adding to each well 0.4 µg of purified plasmid or bacmid, precipitated with 4 µl DOTAP in 500 µl of serum-free medium. For HeLa, 293T and C6/36 cell lines, a mix of 500-µl serum-free medium containing 2 µl of Lipofectamin (Invitrogen) and 0.4 µg of plasmid was used for transfection. After 6 hours of incubation, cells were washed slightly with PBS and fresh serum-supplemented medium was added. Total RNA was isolated from C6/36 and LD652 cells 48 hours post-transfection (p.t.), using the Stratagene Absolutely RNA miniprep kit (cat #400800, Agilent) according to the manufacturer's instructions, but with the optional Turbo-DNase digestion/DNA removal (Ambion). HeLa, 293T, LD652 and C6/36 cells were used for luciferase assays as described below. The quality of the isolated RNA was assessed by checking the 18S- and 28S-rRNA bands after electrophoresis and the concentration of RNA was estimated by spectrophotometry.

2.5. Baculovirus infection for RNA isolation

Infection and harvesting of baculovirus containing PstDNV coding sequence in Sf9 cells were performed as in the manufacturer's protocol (Invitrogen). However, for our purpose of viral RNA isolation, we usually extracted viral RNA after cell infection with second virus stock at 20 to 30 hours p.i. (before lytic phase). Cell culture medium was removed by quick centrifugation and

washed to remove all trace of medium before RNA extraction. Total RNA isolation was performed using the Stratagene Asolutely RNA miniprep kit as described.

2.6. Northern blotting

^{32}P-labeled RNA probes targeting the overlapping domains of the NS1 and NS2 ORFs or the VP ORF were prepared using primers given in Table 3-1 and 3-2 (Publication 1 and 2, Section 1 in Results). First, PCR was performed and the amplified products were then used as templates for transcription *in vitro* with ^{32}P-labeled UTP and T7 RNA polymerase (Cat #M0251S, NEB). Northern blots were done using 10 to 20 µg of total RNA depending on the abundance of viral mRNAs in different cells or systems. Total RNA was mixed with a loading buffer, (50% glycerol, 1 mM EDTA, 4.4% bromophenol blue), denatured at 65^{0}C for 10 minutes and ran on denatured agarose gels (1% agarose) with MOPS buffer and formamide (18 ml of 37% formamide for total 100 ml of MOPS). Gels were then soaked in 20X SSC (3 M sodium chloride and 300 mM trisodium citrate, pH 7.0) to remove formaldehyde as much as possible. The transferring process with positively charged nylon membranes in 20X SSC was carried out for at least 6 hours at RT. Membranes were removed and RNA was fixed by UV-crosslinking. Membrane was stained after this step and photographed to ensure the presence of transferred RNA. The hybridization step was performed overnight, at 42ºC in a solution containing 50% formamide, 5% 20X SSPE (3.0 M NaCl, 0.2 M NaH_2PO_4, and 0.02 M EDTA at pH 7.4.), 10% 50X Denhardt solution (1% Ficoll, 1% polyvinylpyrrolidone, and 1% bovine serum albumin), 1mg/ml salmon sperm DNA, and 0.1% SDS. Following hybridization, membranes were washed with low stringency buffer (2X SSC, 0.1% SDS) and high stringency buffer (0.2X SSC, 0.1% SDS) to remove unhybridized probes. Blots were removed and wrapped in a plastic wrap to prevent drying out and exposed overnight. Blot imaging by Typhoon scanner was described for RNase Protection Assay (RPA).

2.7. Transcript mapping by 5'- and 3'-RACE

The RACE system was used to characterize the 5'-starts and 3'-ends of the polyadenylated transcripts with the FirstChoice® RLM RACE kit (cat #AM1700, Invitrogen) according to the instructions of the supplier. In short, for 5'-starts determination, contaminating RNA or DNA was dephosphorylated with alkaline phosphatase. The caps of the mRNA were removed by tobacco acid pyrophosphatase treatment to leave mRNAs with a single 5'-terminal

phosphate. 45-base RNA 5'-adapter, provided in the kit, is next ligated to the decapped phosphorylated RNA using T4 RNA ligase. Reverse transcription was initiated with hexamer random primers. A PCR was performed first with the 5'-outer adapter from the kit and a specific primer for virus. A nested PCR was then achieved using the 5'-inner adapter primer and nested specific primers of viral sequences. Essentially the same approach, without the phosphatase treatments, was taken for the 3'-RACE. There are many A- stretches in the PstDNV and AalDNV genome and that could lead to mispriming, especially when establishing the 3'-ends. These A-stretches could be confused with poly(A)-tails. The use of anchored antisense T-adapters (Adapter sequence-(T)24 -NV) was essential and PCRs using primers just inboard or outboard of the putative 3'-ends were performed to confirm these end positions.

2.8. RNase protection assays (RPAs)

RPA probes for NS and VP regions to detect NS and VP transcripts were prepared by PCR (Table 3-2, Publication 2, Section 1 in Results), covering the areas of the putative 5'-initiation of the AalDNV transcripts, using *in vitro* transcription by T7 RNA polymerase and ^{32}P-labeled UTP, as described above for Northern blot probes. For a positive control reaction to each predicted site, a sense RNA was generated by transcription *in vitro* and RPA was performed in parallel. All generated probes and control RNA were purified from polyacrylamide gel and used in the RPA III™ Ribonuclease Protection Assay kit (Cat #AM1415, Invitrogen). All reactions were run on denaturing polyacrylamide gels and exposed to a phosphor screen. The screen was then removed and scanned by the Typhoon phosphorimager with ImageQuant software.

2.9. Luciferase assays

For luciferase assays, all viral promoter construct were sequenced (Sanger's method) to confirm sequence and orientation of the insert. Cells were harvested 40 hours p.t. and resuspended in 100 µl of Bright-Glo lysis buffer (Promega). Cell lysates were quickly centrifuged to remove cell debris. 25 µl of the cell extract was used to determine relative luciferase activity according to the Bright-Glo™ Luciferase Assay System's instructions (Cat #E2160, Promega) by the Lumat LB 9507 (Berthold) and compared to background activity in extracts from cells transfected with pGL4 without promoter inserts.

3. Results

3.1. Section 1: Expression strategy of two brevidensoviruses, (*Lipo*)*Penaeus stylorostris* densovirus (PstDNV) and *Aedes albopictus* densovirus (AalDNV)

3.1.1. Résumé en français des différents travaux sur les modalités d'expression de deux brevidensovirus, le densovirus de la crevette (*Lipo*)*Penaeus stylirostris* (PstDNV) et le densovirus du moustique *Aedes albopictus* (AalDNV).

Introduction

Sur la base des ressemblances de leur génome: ADN linéaire simple brin de petite taille (environ 4 kb), organisation comparable des gènes sur un seul brin (génome monosens), absence du motif PLA2 dans la région N-terminale de VP, l'AalDNV et le PstDNV ont été classés à l'origine dans le genre *Brevidensovirus* (le PstDNV comme «membre possible») de la sous-famille des *Densovirinae*. Les résultats de l'analyse de leurs stratégies d'expression respectives, présentés dans les deux publications ci-dessous, montrent qu'en fait ils diffèrent de façon significative. De plus, alors que l'AalDNV possède, comme tous les parvovirus à ses extrémités 5' et 3' une structure de type «épingle à cheveux» essentielle pour la réplication, aucune structure de ce type n'a été rapportée dans les nombreuses séquences d'isolats de PstDNV déposées dans GenBank. Nous avons donc réalisé le clonage et le séquençage complet du génome d'un PstDNV d'origine vietnamienne afin de caractériser ses extrémités.

Publication 1: Hanh T. Pham, Françoise-Xavière Jousset, Jonathan Perreault, Hiroko Shike, Jozsef Szelei, Max Bergoin, and Peter Tijssen, à resoumettre en 2014. **Séquence complète et stratégie d'expression du densovirus de la crevette (*Lipo*)*Penaeus stylirostris* (PstDNV).**

L'instabilité, lors du clonage, des génomes des densovirus est bien connue et très souvent leurs séquences 5' et 3'-terminales riches en GC sont incomplètes. En vue d'obtenir la séquence la plus complète possible d'un densovirus de crevettes, nous avons d'abord cloné les extrémités 5' et 3' du génome de l'isolat VN07 du PstDNV d'origine vietnamienne dans le vecteur linéaire pJAZZ (Lucigen Corp), adapté au clonage de séquences instables, avant de le transférer dans un vecteur circulaire. La région centrale a été également clonée après amplification par PCR. Le séquençage du génome complet (3912 nt) a confirmé l'organisation de type monosens de la région centrale avec deux ORFs codant pour les protéines non-structurales NS1 et NS2 (cette dernière étant localisée entièrement dans la séquence de NS1 mais dans un autre cadre de lecture) et un ORF de petite taille codant pour les trois protéines

structurales VP1, VP2 et VP3. Le séquençage a également confirmé l'absence de structures terminales de type épingle à cheveux, mais a surtout révélé la présence à chaque extrémité d'une longue répétition directe (DTR : Direct Terminal Repeat) de 81 nucléotides. Cette structure se retrouve, en partie seulement, dans certaines séquences de PstDNV déjà publiées mais n'avait jamais été décrite chez les parvovirus. En utilisant des couples d'amorces situées l'une en amont de la DTR 3'-terminale du brin (+), l'autre en amont de la DTR 3' terminale du brin (-), nous avons pu démontrer la présence de structures concatémériques résultant de l'appariement de deux chaînes complémentaires par leurs séquences DTR 3'-terminales. Ces appariements permettent de proposer un mode de réplication de type «cercle roulant» typique des parvovirus et des circovirus. Si ce type de réplication est expérimentalement confirmé, il devrait conduire à la création d'un nouveau genre de densovirus.

L'analyse des transcrits du PstDNV par Northern blot et la technique «RACE» a révélé l'existence de trois transcrits co-terminant dans la région 3' terminale du génome, en aval du cadre de lecture des VP. Deux transcrits d'environ 3.4 kb, l'un sous contrôle du promoteur P2 et épissé dans sa région 5'-terminale d'une séquence intronique de 136 nt, et l'autre sous contrôle du promoteur P12 codent respectivement pour NS1 et NS2. Le transcrit VP d'environ 1.4 kb est sous le contrôle du promoteur P61. Ces résultats permettent de corriger certaines données erronées publiées précédemment faisant état de 3 transcrits pour NS1 et 2 transcrits pour NS2 (Dhar *et al.*, 2010) se terminant tous dans des régions de la séquence codante de NS1 très riches en A. Ces erreurs résultent vraisemblablement de l'appariement non spécifique dans ces régions de l'amorce oligo-dT servant à l'amplification des régions terminales des ARNs messagers. La fonctionnalité des 3 promoteurs a été démontrée par l'insertion de leurs séquences dans le plasmide pGL4.20 (Promega) en amont de la séquence codante de la luciférase et par la transfection avec ces constructions de la lignée cellulaire d'insecte Ld652 et de cellules humaines HeLa. Les trois promoteurs permettent l'expression de la luciférase dans les deux types de cellules.

Publication 2 : Pham HT, Jousset FX, Perreault J, Shike H, Szelei J, Bergoin M, Tijssen P., J. Virol. 2013, Sep; 87: 9928-9932. **Stratégie d'expression du densovirus du moustique *Aedes albopictus*.**

L'analyse des transcrits de l'AalDNV a été réalisée à partir d'ARN extrait de cellules C6/36 transfectées par un plasmide contenant la séquence complète du génome. Elle a été abordée par plusieurs techniques : Northern blot, RACE, cartographie d'amplicons et protection à la RNase. Deux transcrits ont été détectés par Northern blot, l'un de 3.4 kb correspondant aux transcrits NS, l'autre de 1.4 kb correspondant aux transcrits VP. L'analyse des points de démarrage de transcription par la technique RACE a montré l'existence de deux points de démarrage de transcription pour les NS très proches l'un de l'autre et situés de part et d'autre de l'AUG de NS1 (position 335). L'un en position 329, correspond au transcrit NS1, l'autre 14 nucléotides en aval (position 343) correspond au transcrit NS2. Ces transcrits sont sous le contrôle des deux promoteurs P7 et P7.4 se chevauchant. Le point de démarrage de transcription du messager VP a été localisé à la position 2441, soit à 158 nucléotides en amont de l'AUG de VP1. Ce transcrit est sous le contrôle du promoteur P61. Ces résultats ont été confirmés par la technique de protection à la RNase. Les deux transcrits NS et le transcrit VP co-terminent en position 3662, en aval de la séquence codante VP.

La fonctionnalité des promoteurs P7-P7.4 et P61 a été démontrée en clonant leur séquence dans le vecteur d'expression pGL4 (Promega) en amont du gène luciférase et en transfectant des cellules Ld652 et HeLa avec ces constructions. La mesure de l'expression de la luciférase 40h post-transfection a montré que les deux promoteurs NS et le promoteur VP étaient fonctionnels aussi bien en cellules d'insectes qu'en cellules humaines. Une analyse détaillée de l'activité de chacun des deux promoteurs NS P7/P7.4 a été réalisée en clonant séparément les différents éléments: TATA-box, Inr et DPE et en effectuant des mutations, des «knockout», ou des insertion de codon stop en aval de l'ATG de NS1 ou NS2. Après transfert dans le plasmide pGL4, l'activité des différents constructions a été mesurée. Ces analyses ont permis de confirmer la fonctionnalité de chacun des deux promoteurs P7 et P7.4, mais de démontrer également qu'en absence de synthèse du transcrit NS2, le transcrit NS1 contribue de façon significative à l'expression de NS2 par «leaky scanning». L'ensemble de ces données montre que la stratégie d'expression de l'AalDNV diffère de façon significative de celle utilisée par le PstDNV.

3.1.2. Contribution of authors to Publications in Section 1

Françoise-Xavière Jousset and Hiroko Shike were responsible for making PstDNV clone and started the project. Jozsef Szelei provided some initial technical advice. Hanh T. Pham was responsible for conducting all the experiments, analyzing data, and co-responsible for planning. Jonathan Perreault was supervisor for all the works related to RNase Protection Assays. Hanh Pham and Peter Tijssen prepared the manuscripts. Peter Tijssen and Max Bergoin planned and supervised the project.

3.1.3. Publication 1: Complete sequence and expression strategy of (*Lipo*)*Penaeus stylirostris* densovirus (PstDNV)

(To be submited to JGV in 2014)

Hanh T. Pham[1], Françoise-Xavière Jousset[2], Jonathan Perreault[1], Hiroko Shike[3], Jozsef Szelei[1], Max Bergoin[1], and Peter Tijssen[1]

[1] INRS-Institut Armand-Frappier, Laval, Québec, H7V 1B7, Canada

[2] Laboratoire de Pathologie Comparée, Université Montpellier II, 34095 France

[3] Division of Clinical Pathology, HLA Laboratory H160, Milton S. Hershey Medical Center Hershey, Pennsylvania State University, Hershey, PA 17033-0850

Abstract

The shrimp (*Lipo*)*Penaeus stylirostris* densovirus (PstDNV) is possibly related to brevidensoviruses of mosquitoes. We demonstrated that its 3912-nts ssDNA has direct terminal repeats of 81 nts enabling its replication. Transcription maps obtained by Northern blotting, RACE analysis, and amplicon mapping were unlike those of brevidensoviruses. Three promoters resulting in co-terminating mRNA for NS1 (spliced), NS2, and VP, respectively controlled the transcription of PstDNV. Promoter activities were explored using luciferase assays.

Parvoviruses infecting invertebrates are classified in the subfamily *Densovirinae* and subdivided into four genera *Densovirus*, *Pefudensovirus*, *Iteravirus*, and *Brevidensovirus* although some densoviruses (DNVs) do not fit in any of these genera (1-3).

PstDNV (IHHNV) is a notorious shrimp pathogen (4-6) and has been suggested to be related to brevidensoviruses (7, 8). However, phylogenetic relatedness is rather low and ICTV has listed it as a related virus but not approved as a species of the genus *Brevidensovirus* (1). The left, mid, and right ORFs of the PstDNV sequence (3873 nts, ~98% complete; GenBank accession no. AF273215; (7)) have potential coding capacities of 666 amino acids (aa), 363 aa, and 329 aa, respectively. The left 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. Expression of the right ORF yielded virus-like particles (VLPs), the structure of which was resolved to 2.5 Å by X-ray crystallography (9). It is composed of 60 units of 37.5 kDa-VPs with T=1 symmetry. Another PstDNV sequence was deposited in GenBank (AF218266, Hawaiian isolate) with a length of 3909 nts. Finally, sequences obtained after PCR amplifications of other isolates with primers, based on the Hawaiian isolate, were also assumed complete (10, 11).

Terminal hairpin structures involved in parvovirus replication, have so far not been reported for PstDNVs. Circular intermediates have been observed for other parvoviruses (12-14) but, subsequently, hairpins were identified for them (15, 16). The published PstDNV sequences have different extensions at their termini (Figure 3-1A) often through PCR-artifacts (e.g. AF218266.1: 4075 nts, AF218266.2: 3909 nts, and AF273215: 3873 nts). With the approaches taken for PiDNV (17) and JcDNV (18), the ends of the genome of a Vietnamese PstDNV isolate (VN07) were cloned while the central part of the genome was amplified by PCR. The first striking observation was that its sequence, obtained with Sanger's method, contained long direct terminal repeats (DTRs); previously reported genomes missed at least part of their DTRs or contained PCR-generated DTR artifacts. These genomes could readily be used for replication (Figure 3-1B) such as by "rolling-circle" replication for parvoviruses (19) and circoviruses (20). The existence of these concatemeric replication intermediates was confirmed by PCR (Figure 3-1C). The GenBank accession number of the VN07 isolate of PstDNV is KF031144.1 (3910 nts) and has, like the Taiwan B isolate, a deletion of 2 nts in the NS1 intron compared to most GenBank entries.

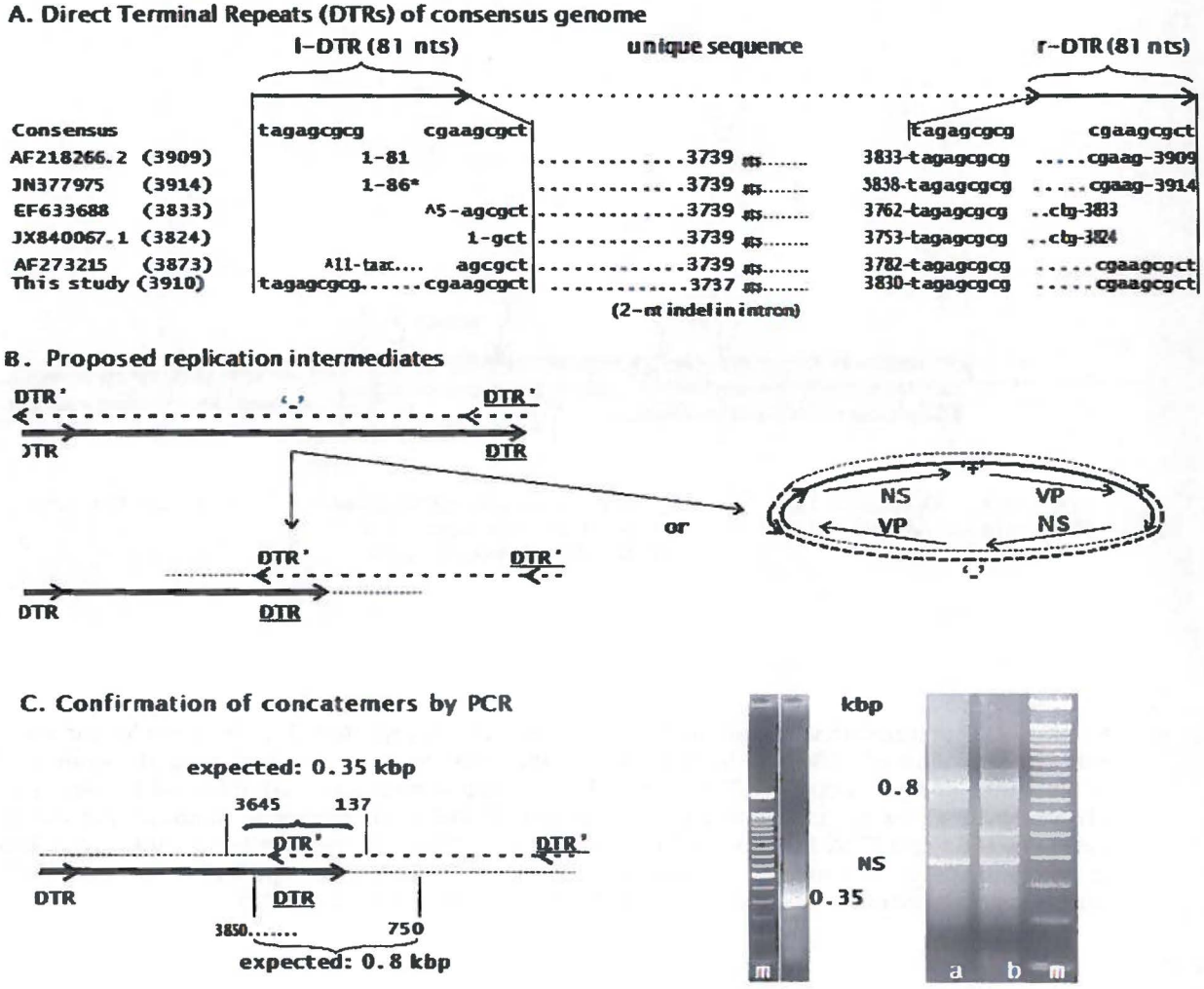


Figure 3-1. (A). PstDNVs with the longest reported genomes (GenBank accession numbers and total length) display incomplete termini. The combination of these sequences reveal a consensus 3912-nt genome with direct terminal repeats (DTR) of 81-nts.The existence of these DTRs has been confirmed by direct cloning and sequencing in this study. The 5-nt insertion between nt 75 and 76 in the l-DTR of JN377975 is probably a sequencing error since it is not matched in the r-DTR (^, first 10 nts in AF273215 seem also incorrect).

(B). It is possible to anneal the direct repeats of the (+) and (-) strands to create linear or circular concatemers and fill in (fine-dashed lines) and typical "rolling-circle" DNA replication (19, 20).

(C). The existence of these concatemers was confirmed by PCR with extracted DNA from purified virions as template. Two sets of primers were used to generate amplicons across the termini (agarose gel electrophoresis shown, a and b are two different virus purifications), followed by confirmation by semi-nested PCR and sequencing (NS, nonspecific).

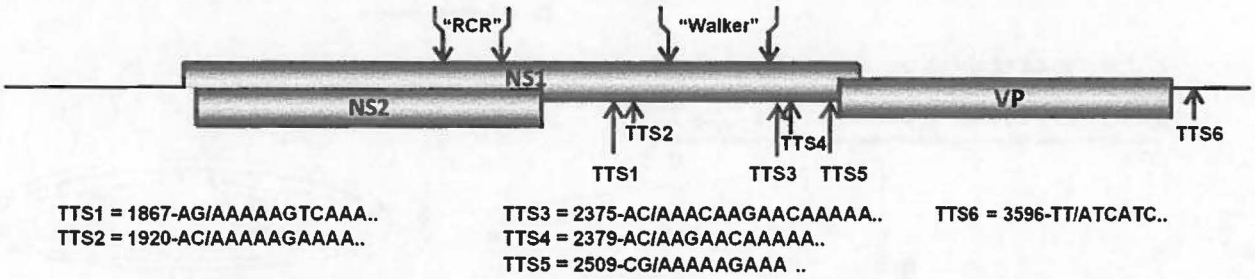


Figure 3-2. Transcription termination sites (TTSs) previously reported (21). The poly(A)-tails are all, with the exception of TTS6, within ORFs and would result in the addition of long poly-lysine tails to the translation products. The TTS in the ORFs, indicated by a slash, are followed by very A-rich sequences that are prone to false priming during 3'-RACE experiments. Although the authors assigned TTS1 and TTS2 to the left ORF transcript and TTS3-5 to the middle ORF transcript, there is no evidence to support this, since these transcripts were not separately analyzed. (RCR: Rolling-circle replication motif, Walker = Walker/helicase/NTP-binding motif).

It has been reported that PstDNV generates three transcripts for NS1, two for NS2 and one for VP under the control of P2, P11 and P61 promoters, respectively (21, 22) (Figure 3-2). However, those results (21) were dubious since: (i) its Northern blots showed a smear from 0.1 to 4 kb (full genome length) indicative of poor-quality mRNA and contaminating genomic DNA; (ii) a mixture of mRNAs was analyzed and termination sites can thus not been assigned to a particular PstDNV mRNA; (iii) all reported NS transcript termination sites are within the NS1 ORF (Figure 3-2) and lack the canonical termination motifs; (iv) the positions of reported poly(A) sites are in A-rich sequences (Figure 3-2) readily leading to nonspecific priming in the 3'RACE (the A-content of the PstDNV genome is 36%); (v) the reported poly(A) of the NS transcripts were in frame within the NS1-ORF and would lead to the addition of long lysine tails in translation products; and (vi) NS1 would be truncated near, or upstream of the essential NTPase motif. Since this is unlikely, we examined the transcript mapping and promoter activity of PstDNV.

A sequence of 3873 nts of PstDNV (AF273215) was subcloned in the pCR2.1 vector (Invitrogen) at EcoRI sites. As in the BacMam system, this PstDNV insert was also transferred into a baculovirus, via the pFastbac1 baculovirus expression vector (Invitrogen), from which the polyhedrin promoter (nt 3904-4032) was removed by introducing a BamHI site into pFastBac1 at position 3870 and digestion with BamHI at positions 3870 and 4032. The PstDNV genome was then subcloned between BamHI and XhoI sites (Pst-baculovirus). C6/36 cells, in which shrimp HePV parvovirus can be propagated (23), LD652 and HeLa cells were chosen for transfection using Lipofectamin (Invitrogen protocol). C6/36 cells were grown in RPMI medium (HeLa cells in DMEM and LD652 in Sf900 medium), supplemented with 5% foetal bovine serum (FBS). After 48 hrs, RNA was isolated using the Stratagene Absolutely RNA miniprep kit. For Northern blots, PCR-amplified products (primers in Table 3-1), that targeted overlapping domains of the NS1 and NS2 ORFs or VP ORF, were transcribed *in vitro* with [α- ^{32}P]UTP and T7 RNA polymerase (NEB) (24). With 20 µg of total RNA (24, 25) from Pst-baculovirus-transfected cells, transcript sizes of 3.4 and 1.2 kb were revealed with the VP-specific probe and a 3.4 kb transcript with the NS-specific probe (Figure 3-3A). The 5'-initiation and 3'-termination of PstDNV-mRNAs from pCR2.1-PstDNV-transfected C6/36 cells were identified with the FirstChoice™ RLM RACE kit (Invitrogen). The 5' adapter (IP) and the 3' anchored (AP) primers (in the kit) were used in PCR with gene-specific primers (Table 3-1) as shown in Figure 3-4. These amplicons were cloned in pGEM-T and sequenced. The ~500-bp amplicon revealed that NS1 transcription started at nt position 100 (CAGTTT), 58 nt upstream of NS1-AUG initiation codon (Figure 3-4A; Figure 3-5A, numbering as in AF273215). There was an intron of 135 nts

between nt 287 to 423 (Figure 3-4A; Figure 3-5D). Sequencing of the ~250 bp amplicon revealed that NS2 started at nt 465 (ACATTC) (Figure 3-4A; Figure 3-5B), i.e. 77 nt upstream of the AUG-NS2 initiation codon. Sequencing of amplicons obtained with the P2622R and P2542R primers located the VP mRNA start at nt 2442 (CAGTACG), 99 nts upstream of the initiation of translation (nt 2541) (Figure 3-4A; Figure 3-5C). Primers P3182F and P2600F (just downstream of the start of the VP transcription) used with the AP primer were designed for characterization of 3'polyadenylation sites of viral transcripts using the 3' RACE method (Figure 3-4B). Sequencing of these amplicons showed that the transcripts terminated at nt 3592 (Figure 3-5E), 15 nts downstream of the last polyadenylation signal, at nt 3577-3582. In order to confirm that the NS transcripts co-terminated with the VP transcript, primer P2347F, upstream of the VP transcript start, and two primers, P3545R upstream of poly(A) and P3640R downstream of the poly(A) site, were used to obtain amplicons by reverse transcription and PCR (Figure 3-3B). Amplicons were obtained only with the reverse primer P3545R (Figure 3-3C) indicating that NS and VP transcripts used the same polyadenylation site.

Promoter regions were amplified by PCR (primers in Table 3-1) and cloned into EcoRV in pGL4.20, upstream of the luciferase gene (Promega, GenBank DQ188840) to estimate their functionality. Sequencing confirmed insert orientation. Those with reverse promoter orientation served as controls. The MIDNV NS promoter (insect virus replicating in LD652 cells; (26)) and the SV40 promoter were also cloned into pGL4.20 vectors to serve as positive controls. Transfection of LD652 using DOTAP (Roche) and HeLa cells using Lipofectamin (Invitrogen) in 6-well plates was according to suppliers' instructions. Cells were harvested 40 hours post-transfection and relative luciferase activity was determined according to Promega's instructions. The PstDNV promoters were functional both in insect and human cells, in contrast to constructs lacking a promoter or promoters in reverse orientation (data not shown), whereas the SV40 promoter preferred HeLa cells and the MIDNV promoter preferred insect cells (Figure 3-3D).

In conclusion, PstDNV demonstrated, in addition to phylogenic divergences, significant differences, such as DTRs vs hairpins and expression strategy (number of promoters, introns), with brevidensoviruses, despite superficial similarities (genome size, ORF structure, transcript sizes). PstDNV probably uses DTRs for their replication, in contrast to (other) parvoviruses that have long terminal hairpins. Finally, all PstDNV transcripts co-terminate downstream of the VP gene.

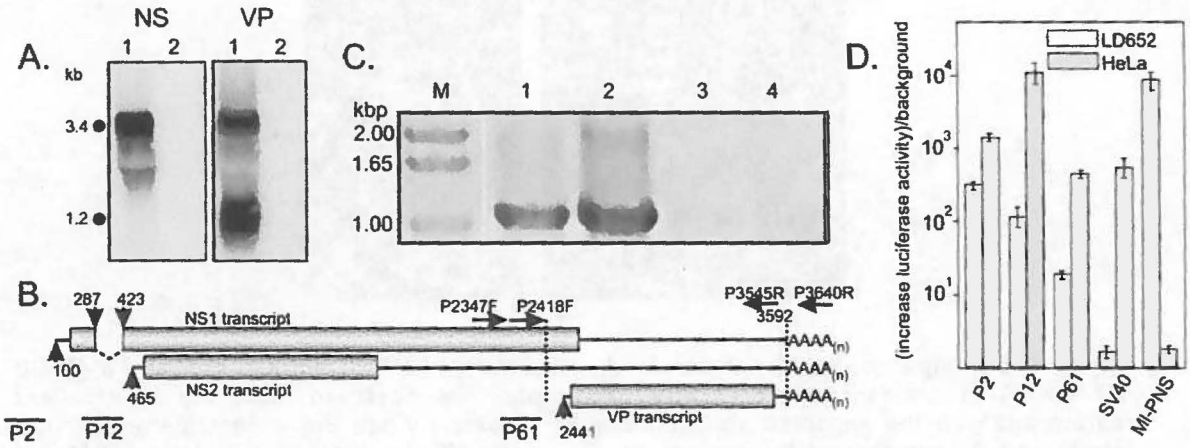


Figure 3-3. (A). Northern blots with NS and VP probes (lane 1: PstDNV-transfected cells and lane 2: mock-transfected cells).(B). Transcription map obtained with RACE confirmed results obtained with Northern blotting. To confirm that the NS and VP transcripts co-terminate, a semi-nested PCR, first with P2347F and then with P2418F, was designed with the 4 primers indicated by arrows (amplicon mapping; number is 5'-nt in primer and F or R forward and reverse). (C). M indicates size markers; 1, P2347F/P3545R; 2, P2418F/P3545R (with product from lane 1); 3, P2347F/P3640R; 4, P2418F/P3640R (with product from lane 3). These results were also in agreement with the Northern blot and RACE experiments. (D). The PstDNV promoters were active in both insect and vertebrate cells.

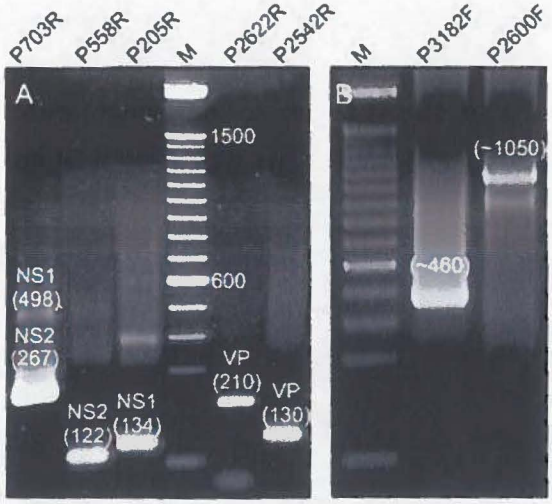


Figure 3-4. Analysis of amplicons obtained with different PstDNV primers in 5' and 3'-RACE (M is 100 nt-size marker) of PstDNV transcripts. The observed sizes are in excellent agreement with the expected sizes (values in brackets) when the expected transcription starts and the position of the primers on the ligated 5'-oligonucleotide and the cDNA are taken into account. (A. 5'-RACE of NS and VP transcripts; B. 3'-RACE NS and VP transcripts).

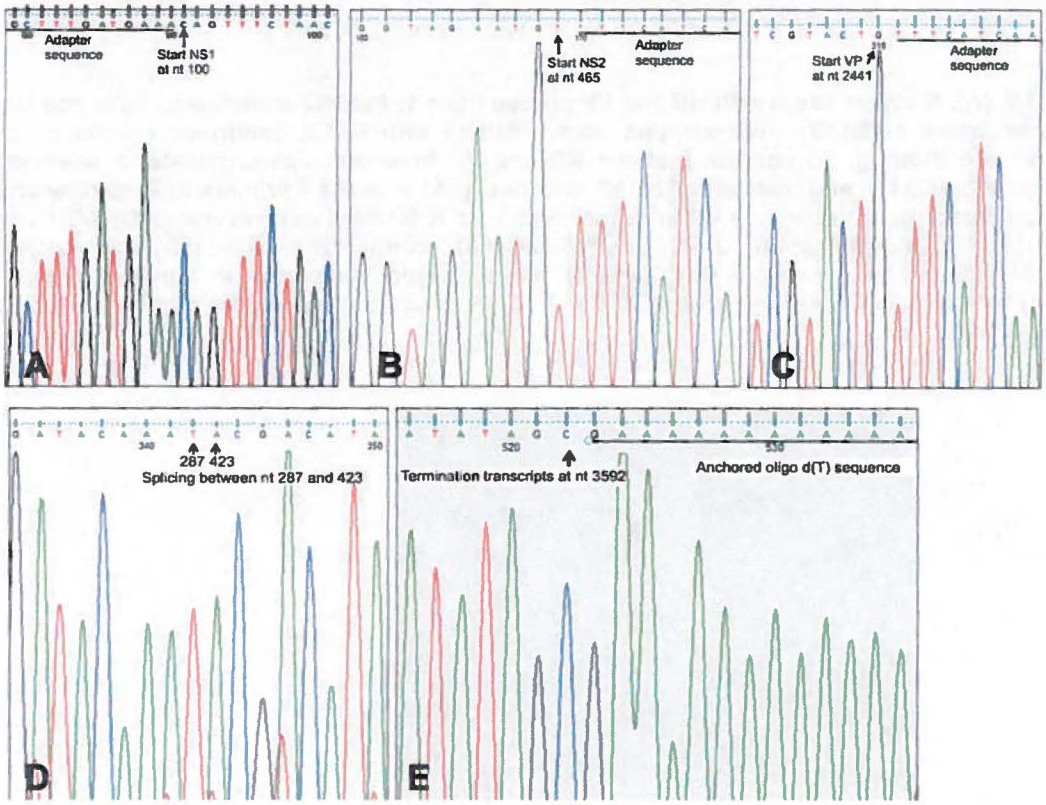


Figure 3-5. Sequencing results of splicing site, 5' and 3'-RACE. A. Start NS1 transcription. B. Start NS2 transcription. C. Start VP transcription. D. Splicing in NS1 transcript. E. Termination of NS and VP transcripts.

Table 3-1. Primers for RACE, amplicon mapping and promoter amplicons

Primer names	Sequences	Purpose
P2600F	TGAAGACGAAGAACACGCCGAAG	3'-RACE
P3182F	TGGTGACACTAGAAATTGGTATG	3'-RACE
P205R	TGACGGACTAGGTATTGATTGC	5'-RACE of NS
P558R	CACTGTCCGTTGACATTG	5'-RACE of NS
P703R	TGGAGAAATTCCCTGGCTGG	5'-RACE of NS
P2542R	ATGGTTGTCTATGATGTC	5'-RACE of VP
P2622R	CTTCGGCGTGTTCTTCGTCTTC	5'-RACE of VP
P2418F	GCCTCTCAGACAGGATGAAC	Amplicon mapping
P3545R	GCACATCGAATACATTTAG	Amplicon mapping
P3640R	TTGCCTGGGTAGCTGGTATG	Amplicon mapping
5'-adapter	GCUGAUGGCGAUGAAUGAACACUG CGUUUGCUGGCUUUGAUGAAA	5'-RACE Adapter Ambion
3'-adapter	GCGAGCACAGAATTAATACGACTCA CTATAGGT12VN	3'-RACE Adapter Ambion
5'-Outer	GCTGATGGCGATGAATGAACACTG	5'-RACE Outer Ambion 2nd round
5'-Inner	CGCGGATCCGAACACTGCGTTTGC TGGCTTTGATG	5'-RACE Inner Ambion 2nd round
3'-Outer	GCGAGCACAGAATTAATACGACT	3'-RACE Outer Ambion 2nd round
3'-Inner	CGCGGATCCGAATTAATACGACTCA CTATAGG	3'-RACE Inner Ambion 2nd round
P21F	TGCGAGCGCTTCGCAGAAACC	P2 promoter PstDNV
P157R	GTTCCAAAATTGTCCTTAGTC	P2 promoter PstDNV
P289F	A AGTACAAGTGACTGACTAAG	P12 promoter PstDNV
P542R	TGAGATTGTCTATAAAACACTCG	P12 promoter PstDNV
P2347F	GGTACCTCCAGCTGATGGT	P61 promoter PstDNV
P2540R	GGTTGTCTATGATGTCGTCG	P61 promoter PstDNV

*Numbers in names of primers refer to their 5'-nucleotide position in AF273215

References

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.

2. 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.

3. Bergoin M, Tijssen P. 2010. Densoviruses: a highly diverse group of arthropod parvoviruses, p. 57-90. In Asgari S, Johnson KN (ed.), Insect Virology. Horizon Scientific Press, Norwich, U.K.

4. Flegel TW. 2006. Shrimp parvoviruses, p. 487-493. In Kerr JR, Cotmore SF, Bloom ME, Linden RM, Parrish CR (ed.), Parvoviruses. Hodder Arnold, London.

5. Chayaburakul K, Lightner DV, Sriurairattana S, Nelson KT, Withyachumnarnkul B. 2005. Different responses to infectious hypodermal and hematopoietic necrosis virus (IHHNV) in *Penaeus monodon* and *P. vannamei*. Dis Aquat Organ 67:191-200.

6. Lightner DV, Redman RM, Bell TA. 1983. Infectious hypodermal and hematopoietic necrosis, a newly recognized virus disease of penaeid shrimp. J Invertebr Pathol 42:62-70.

7. Shike H, Dhar AK, Burns JC, Shimizu C, Jousset FX, Klimpel KR, Bergoin M. 2000. Infectioushypodermal and hematopoietic necrosis virus of shrimp is related to mosquito brevidensoviruses. Virology 277:167-177.

8. Roekring S, Nielsen L, Owens L, Pattanakitsakul SN, Malasit P, Flegel TW. 2002. Comparison of penaeid shrimp and insect parvoviruses suggests that viral transfers may occur between two distantly related arthropod groups. Virus Res 87:79-87.

9. Kaufmann B, Bowman VD, Li Y, Szelei J, Waddell PJ, Tijssen P, Rossmann MG. 2010. Structure of *Penaeus stylirostris* densovirus, a shrimp pathogen. J Virol 84:11289-11296.

10. Rai P, Safeena MP, Karunasagar I. 2011. Complete nucleic acid sequence of *Penaeus stylirostris* densovirus (PstDNV) from India. Virus Res 158:37-45.

11. Kim JH, Kim HK, Nguyen VG, Park BK, Choresca CH, Shin SP, Han JE, Jun JW, Park SC. 2012. Genomic sequence of infectious hypodermal and hematopoietic necrosis virus (IHHNV) KLV-2010-01 originating from the first Korean outbreak in cultured Litopenaeus vannamei. Archives of virology 157:369–373.

12. Lusebrink J, Schildgen V, Tillmann RL, Wittleben F, Bohmer A, Muller A, Schildgen O. 2011. Detection of head-to-tail DNA sequences of human bocavirus in clinical samples. PloS one 6:e19457.

13. Lusebrink J, Wittleben F, Schildgen V, Schildgen O. 2009. Human bocavirus-insights into a newly identified respiratory virus. Viruses 1:3-12.

14. Kapoor A, Hornig M, Asokan A, Williams B, Henriquez JA, Lipkin WI. 2011. Bocavirus episome ininfected human tissue contains non-identical termini. PloS one 6:e21362.

15. Huang Q, Deng X, Yan Z, Cheng F, Luo Y, Shen W, Lei-Butters DC, Chen AY, Li Y, Tang L,Soderlund-Venermo M, Engelhardt JF, Qiu J. 2012. Establishment of a reverse genetics system for studying human bocavirus in human airway epithelia. PLoS Pathog 8:e1002899.

16. Deng X, Yan Z, Luo Y, Xu J, Cheng F, Li Y, Engelhardt JF, Qiu J. 2013. *In vitro* modeling of human bocavirus 1 infection of polarized primary human airway epithelia. J Virol 87:4097-4102.

17. Huynh OT, Pham HT, Yu Q, Tijssen P. 2012. *Pseudoplusia includens* densovirus Genome Organization and Expression Strategy. J Virol 86:13127-13128.

18. Pham HT, Nguyen OTH, Bergoin M, Tijssen P. 2013. Sequence of full-length JcDNV genome. Genome Announc. 2013 Aug 8; 1(4).

19. Cotmore SF, Tattersall P. 2006. A rolling-hairpin strategy: basic mechanisms of DNA replication in the parvoviruses, p. 172-185. In Kerr JR, Cotmore SF, Bloom ME, Linden RM, Parrish CR (ed.), Parvoviruses. Hodder Arnold, London.

20. Faurez F, Dory D, Grasland B, Jestin A. 2009. Replication of porcine circoviruses. Virology journal 6:60.

21. Dhar AK, Kaizer KN, Lakshman DK. 2010. Transcriptional analysis of *Penaeus stylirostris* densovirus genes. Virology 402:112-120.

22. Dhar AK, Kaizer KN, Betz YM, Harvey TN, Lakshman DK. 2011. Identification of the core sequence elements in *Penaeus stylirostris* densovirus promoters. Virus Genes 43:367-375.

23. Madan N, Nambi KS, Abdul Majeed S, Taju G, Sundar Raj N, Farook MA, Vimal S, Sahul Hameed AS. 2013. *In vitro* propagation of hepatopancreatic parvo-like virus (HPV) of shrimp in C6/36 (*Aedes albopictus*) cell line. J Invertebr Pathol 112:229-235.

24. Tijssen P, Li Y, El-Far M, Szelei J, Letarte M, Zadori Z. 2003. Organization and expression strategy of the ambisense genome of densonucleosis virus of *Galleria mellonella*. J Virol 77:10357-10365.

25. 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.

26. Fediere G, El-Far M, Li Y, Bergoin M, Tijssen P. 2004. Expression strategy of densonucleosis virus from *Mythimna loreyi*. Virology 320:181-189.

3.1.4. Publication 2: Expression strategy of *Aedes albopictus* densovirus

Pham HT[1], Jousset FX[2], Perreault J[1], Shike H[3], Szelei J[1], Bergoin M[1], Tijssen P[1]

J Virol. 2013 Sep; 87(17):9928-32

[1] INRS-Institut Armand-Frappier, Laval, Québec, H7V 1B7, Canada

[2] Laboratoire de Pathologie Comparée, Université Montpellier II, 34095 France

[3] Division of Clinical Pathology, HLA Laboratory H160, Milton S. Hershey Medical Center Hershey, Pennsylvania State University, Hershey, PA 17033-0850

Abstract

The transcription map of the *Aedes albopictus* (AalDNV) brevidensovirus was identified by Northern blotting, rapid amplification of cDNA (RACE) analysis and RNase protection assays. AalDNV produced mRNAs (3359 (NS1), 3345 (NS2) and 1246 (VP) nts. The two overlapping P7/7.4 promoters employed closely located alternate transcription initiation sites, positioned at either side of the NS1 initiation codon. All NS mRNAs co-terminated with VP mRNA. All promoters, explored using luciferase assays, were functional in insect and human cell lines.

Introduction

Invertebrate parvoviruses (densoviruses [DNVs]) are subdivided into four genera: *Densovirus*, *Pefudensovirus*, *Iteravirus*, and *Brevidensovirus* (1–3). Brevidensoviruses have a 4.1-kb single-stranded DNA (ssDNA) genome with three open reading frames (ORFs) on the same strand (about 790 [NS1], 360 [NS2], and 350 [VP] amino acids). The genome of *Aedes albopictus* densovirus (AalDNV) (GenBank accession no. NC_004285) has terminal, T-shaped hairpins (4). Brevidensoviruses were isolated from medically important mosquito vectors, such as *Aedes albopictus* (AalDNV) (4), *Aedes aegypti* (AaeDNV) (5), and *Anopheles gambiae* (AgDNV) (6). AalDNV was isolated from *Aedes albopictus* C6/36 cells (7) but is infectious for *Aedes aegypti* larvae (8, 9).

AaeDNV expression has been studied to some extent (10, 11). ORF_{NS1} of AaeDNV was reported to have a 57-amino-acid N-terminal extension compared to that of AalDNV (4, 5). β-Galactosidase fusion proteins with the three ORFs were enzymatically active, except for NS1 (12) unless NS1-β-gal was constructed downstream of the corresponding AalDNV AUG_{NS1} (11). Primer extension demonstrated that AaeDNV VP transcription started at nucleotide (nt) 2402 (10). Here, AalDNV transcription was analyzed by Northern hybridization, 5' and 3' rapid amplification of cDNA ends (RACE), amplicon mapping, and RNase protection assays (RPAs). AalDNV promoter activities were also determined.

AalDNV transcripts were obtained after transfection of pCR2.1-AalDNV (containing the AalDNV genome of 4,176 nt [4] between its EcoRI sites), using Lipofectamine (Invitrogen protocol), in permissive C6/36 cells (7) grown in RPMI medium supplemented with 5% fetal bovine serum (FBS). After 48 h, RNA was isolated using the Stratagene Absolutely RNA miniprep kit.

Northern blotting. ^{32}P-labeled RNA probes for Northern blots targeted overlapping domains of the NS1 and NS2 ORFs or VP ORF. PCR-amplified products (primers in Table 3-2) were transcribed *in vitro* with [α-^{32}P]UTP and T7 RNA polymerase (NEB) (13). Northern blots using 10 g of total RNA (13, 14) revealed transcript sizes of 3.4 and 1.3 kb with the VP-specific probe and a 3.4-kb transcript with the NS-specific probe (Figure 3-6A).

RACE experiments. The FirstChoice RLM RACE kit (Invitrogen) was used to identify the 5' starts and 3' ends of the polyadenylated transcripts. The 5' adapter primer (IP) and the 3' anchored primer (AP) (in the kit) were used in PCR with AalDNV-specific primers (Table 3-2), as shown in Figure 3-6B. Amplicons obtained with A626R/IP (~325 bp), A497R/IP (~200 bp),

A2740R/IP (~350 bp), A2540R/IP (~150 bp), A3162F/AP (~600 bp), and A3461F (~300 bp) (data shown in Figure S1, Annexe 1) were sequenced. Sequencing revealed that NS1 transcription started at nt 329 (AGTA), 6 nt upstream of AUG_{NS1}, and that VP transcription started at nt 2441 (CAGTCG), 158 nt upstream of AUG_{VP} (Figure 3-6B) (sequence data shown in Figure S2, Annexe 1). Sequencing of the 300- and 600-bp amplicons showed a transcription termination position at 3680, 18 nt downstream of the polyadenylation signal at position 3662 (Figure. 3-6C) (sequence data shown in Figure S3, Annexe 1).

As is common for other densoviruses, NS2 could be translated from nt 411 by leaky scanning (13) on the NS1 transcript. The short untranscribed region of NS1 (UTR_{NS1}), in spite of a consensus AnnAUGG sequence for NS1 initiation, could favor it (13). A canonical initiator Inr_{NS2} sequence (CAGT) is located at nt 342. After TfiI digestion (NEB) at nt 338 of the A831R/IP PCR product, followed by a nested PCR (A626R/IP or A497R/IP) for specific amplification of putative NS2 transcripts, bands of 330 and 200 nt were produced (Figure S1-A, Annexe 1). Their sequencing revealed that NS2 transcription started at nt 343 (Figure S2, Annexe 1). Thus, NS1 and NS2 AalDNV transcription starts were separated by 14 nt on either side of AUG_{NS1}.

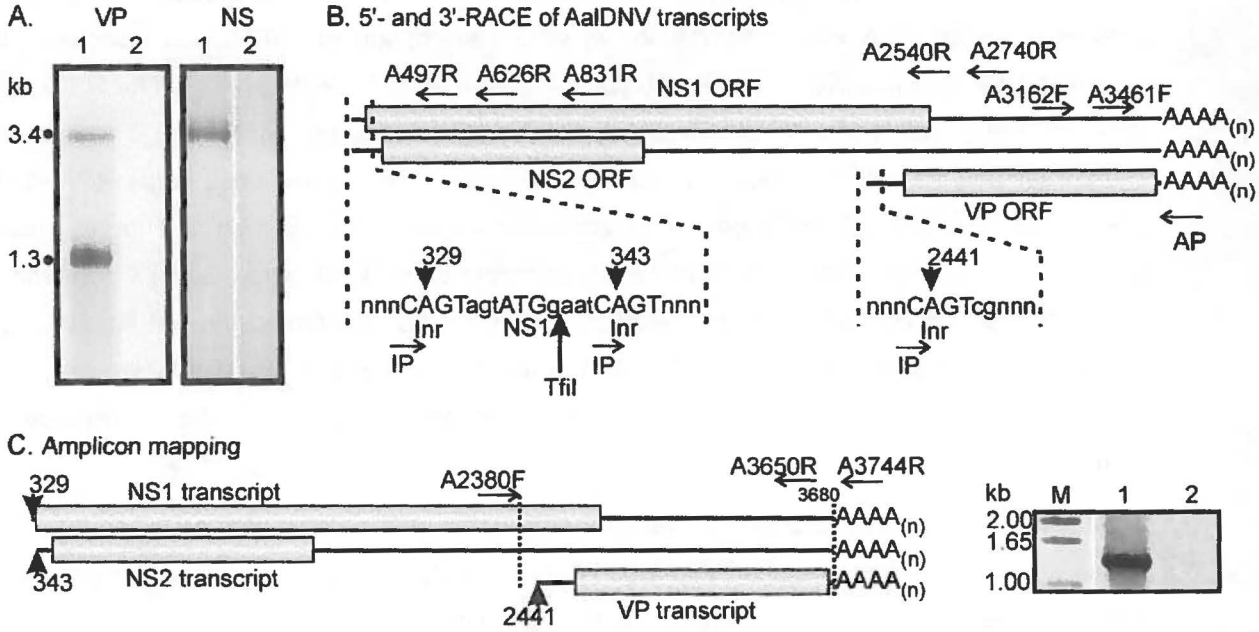


Figure 3-6. (A) Northern blot analysis of AalDNV transcription 48 h after transfection with vectors with virus-specific inserts (lanes 1) or without inserts (lanes 2). The VP probe also detected bands corresponding to the size of NS transcript indicating that VP and NS transcripts may coterminate. (B) Strategy of 5' and 3' RACE. Above the ORFs are the specific primers that were designed, and below are the ORFs and IP and AP primers from the Invitrogen kit (indicated with arrows). We took advantage of the TfiI restriction site between the putative Inr of NS1 and NS2 to distinguish between these transcripts. (After digestion, only amplicons from the second Inr would be obtained). (C) Confirmation of termini of NS transcripts of AalDNV. The diagrams represent transcript map results by 5' and 3' RACE. Forward primers that do not recognize VP transcripts and reverse primers, both upstream and downstream of the VP transcript end (3' RACE), were used in a PCR (indicated in the diagrams). As expected, only the inboard reverse primer (A3650R in lane 1) gave a product. M, markers; lane 2, A2380F/A3744R.

Amplicon mapping. In order to confirm that NS and VP transcripts all coterminated at position 3680, primer A2380F (upstream of the VP transcript) and two reverse primers, A3650R (upstream) and A3744R (downstream), of the VP transcript end were designed for reverse transcription (RT)-PCR amplification (Table 3-2). Only a band of about 1,300 nt was obtained using the A2380F/A3650R set of primers for NS transcripts (Figure 3-6C).

RNase protection assays. RPAs were employed to confirm NS1, NS2, and VP transcription starts. RPA probes for NS and VP transcripts were prepared by PCR (Table 3-2; Figure 3-7A), and *in vitro* transcription as described for Northern blot probes. For positive controls, a sense RNA was generated by *in vitro* transcription and RPA was performed in parallel: for NS, two positive-control RNAs spanning from nt 329 to 439 (predicted for NS1) and nt 343 to 439 (predicted for NS2) were used, and for VP one from nt 2441 to 2542 was used. The probes and control RNA were purified from polyacrylamide gel and used in the RPA III kit (Invitrogen). Size markers were generated according to the 5' RACE results. The NS probe protected the expected 96 and 110 nt of the NS transcript 5' ends (Figure 3-7B), confirming NS1 and NS2 mRNA start positions. Similarly, the VP probe confirmed the VP transcript 5' extremity by protecting 101 nt (Figure 3-7B). A protected band at 219 nts, slightly smaller than the entire probe (236 nts, including 17 nts of unspecific sequence) showed the coexistence of NS transcritps.

Promoter activity. NS (P7/7.4) and VP (P60) promoter regions were amplified by PCR (primers in Table 3-2) and cloned into EcoRV in pGL4.20, upstream of the luciferase gene (Promega; GenBank accession no. DQ188840) to estimate their functionality. Those with the reverse promoter orientation, shown by sequencing, served as controls. The *Mythimna loreyi* (MlDNV) NS promoter (an insect virus replicating in LD652 cells) (15) and the simian virus 40 (SV40) promotes were also cloned into pGL4.20 vectors to serve as positive controls. Transfection of LD652 using DOTAP (Roche) and HeLa cells using Lipofectamine (Invitrogen) in 24-well plates was performed according to the suppliers' in-structions. Cells were harvested 40 h posttransfection and resuspended in 100 l of Bright-Glo lysis buffer (Promega), and relative luciferase activity was determined according to Promega's instructions. The AaIDNV P7/7.4 and P60 promoters were functional both in insect cells and in human cells, in contrast to constructs lacking a promoter or with promoters in the reverse orientation, whereas the SV40 promoter preferred HeLa cells and the MlDNV promoter preferred insect cells (Figure 3-7C).

Table 3-2. Primers used for Northern blot probes, RACE, amplicon mapping and RNase Protection Assays (RPAs)

Primer name	Sequence	Purpose
A639F	GCTCCAGAGCCTCTGAACAGCTTG	NS probe - Northern blot
A1345R	TAATACGACTCACTATAGGGGTTCTGACTC TTGTGCTGTTTC	NS probe - Northern blot
A3106F	CTAGAAACAGTTGCAGCAACCGGAC	VP probe - Northern blot
A3509R	TAATACGACTCACTATAGGCGTACTTGATA TCTGAATTTCATG	VP probe - Northern blot
A3372F	AACTACAACATATGCCACGTCAG	3 -RACE
A3461F	ACAAGTTCCAGACGAAACAGG	3'-RACE
A497R	GTTCGTAATTGTTGGCATTCCT	5'-RACE of NS
A626R	GTGGGTAGATGTTATCAACGG	5'-RACE of NS
A831R	CTTGCCTGTGACCCGTTATTATCC	5'-RACE of NS
A2540R	GTGCGTTGTCTTCTTCTTCTATC	5'–RACE of VP
A2740R	GACCAAACATTACGGAAATGG	5 -RACE of VP
A3126F	CGGACCATTAGCACAACAAAC	3'-RACE
A2380F	GAGTATACAACACAGAGAAG	Amplicon mapping
A3650R	TCATAA GGCATACATGCTAC	Amplicon mapping
A3744R	TCTGTCGTGGACATTATCAG	Amplicon mapping
A272F (+ UTS*)	GCGATGAATGAACACTGAATCCACCACCA CATGATCC	RPA-NS probe
A329F (+T7 sequence)	TAATACGACTCACTATAGGAGTAGTATGGA ATCAG	RPA-Positive control for NS1
A343F (T7 sequence)	TAATACGACTCACTATAGGGTCTGCAGTGA ACATTCG	RPA-Positive control for NS2
A439R	TCTCCTCCTGGATTTACACTG	RPA-Positive control for NS1 and NS2
A439R (+T7 sequence)	TAATACGACTCACTATAGGTCTCCTCCTGG ATTTACACTG	RPA-NS probe
A2323F (+UTS)	GCGATGAATGAACACTGGCATATGAACGA AACCTCAC	RPA-VP probe
A2441F (+T7 sequence)	TAATACGACTCACTATAGGAGTCGGCCACC AGGTCTTGTAG	RPA-Positive control VP
A2542R	ATGTGCGTTGTCTTCTTCTTC	RPA-Positive control VP
A2542R (+T7 sequence)	TAATACGACTCACTATAGGATGTGCGTTGT CTTCTTCTTC	RPA-VP probe
A148F	TCCAATTGGAACACACGGAC	P7/7.4 promoter AalDNV
A333R	CTACTGACTCTCCCTTC	P7/7.4 promoter AalDNV
A2431F	CAAACTCATCAGTCGGCCAC	P60 promoter AalDNV
A2597R	CCTCTGCTTCTTCTTTTGC	P60 promoter AalDNV

*UTS: unspecific-target sequence; R and F: sense and antisense.

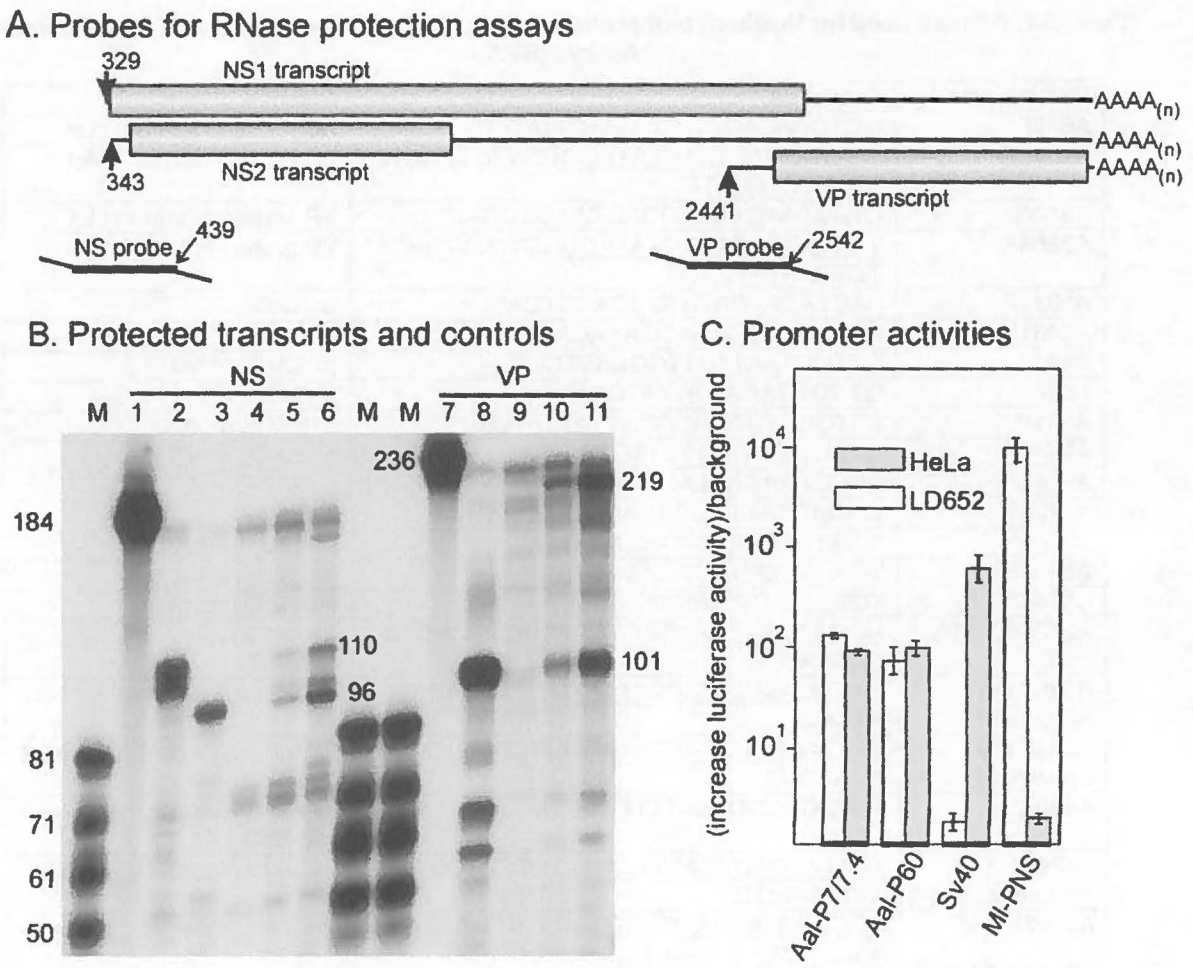
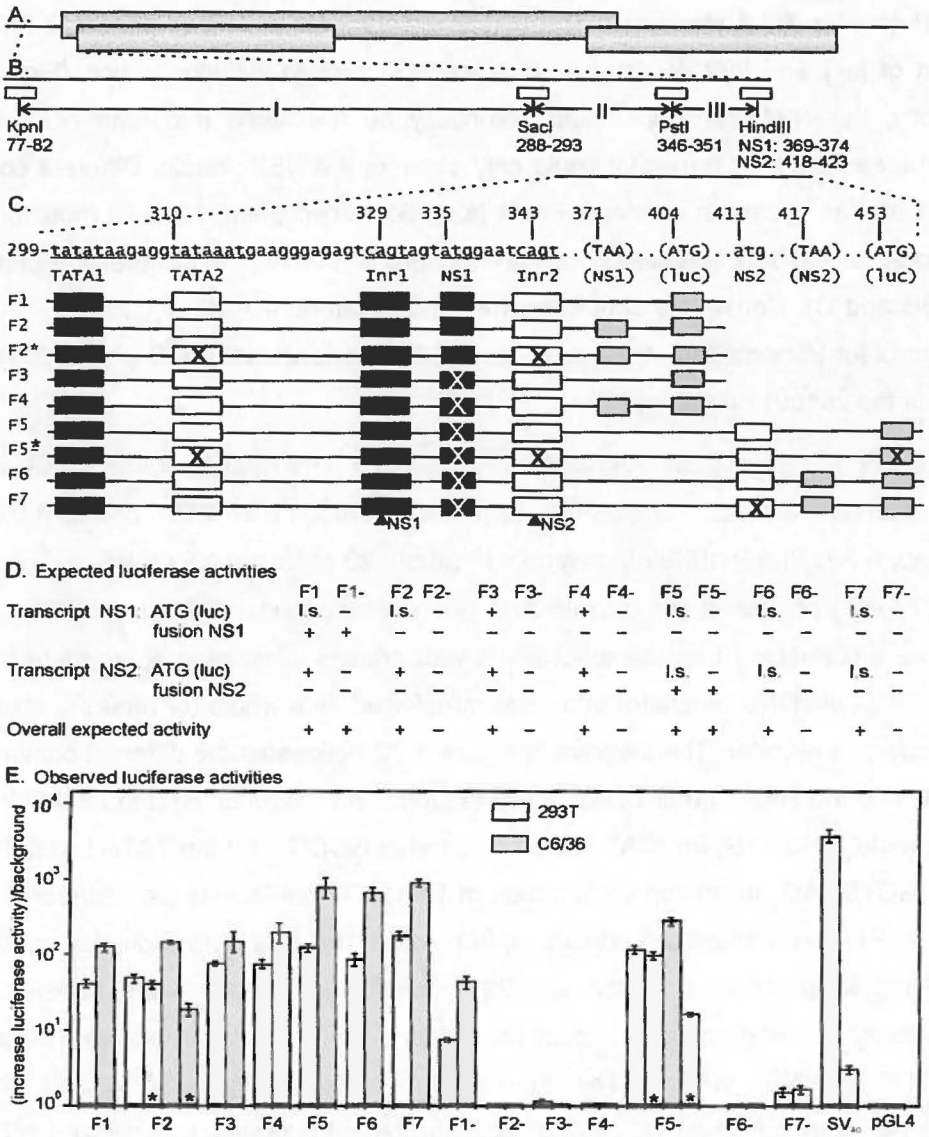


Figure 3-7. (A) Diagram of AalDNV expression and location of probes. (B) RNase protection assay of starts of NS and VP transcripts of AalDNV. Lanes: 1, NS probe with a specific length of 184 nt and short nonspecific extra terminal sequences; 2, positive control for NS1 with specific size obtained with RACE experiments; 3, positive control for NS2 with specific size obtained with RACE experiments; 4, total RNA from C6/36; 5, total RNA from AalDNV-infected C6/36 (15 g); 6, total RNA from AalDNV-infected C6/36 (40 g); 7, VP probe; 8, positive control for VP; 9, total RNA from C6/36; 10, total RNA from AalDNV-infected C6/36 (15 g); 11, total RNA from AalDNV-infected C6/36 (40 g). The band at 101 nt confirmed the VP start, and the band at 219 nt represented protection of the VP probe on the NS transcript. (C) Luciferase activity 40 h after transfection with different promoter constructs as fold increase over background (transfected pGL4 without insert). Inverse orientation of the promoters did not increase activity significantly over background. SV40 (not shown) and the insect virus MIDNV NS promoter (MI-NSP) had a strong preference for cells from vertebrates and invertebrates, respectively. Surprisingly, AalDNV had significant activity in both types of cells.

(i) Promoter 7/7.4 elements. The RACE experiments and RPA (Figure 3-7) revealed the location of Inr1 and Inr2 at either side of the NS1 protein initiation codon (Figure 3-8C). Consequently, the NS1 transcript could potentially be translated into both NS1 and NS2 proteins, whereas the NS2 transcript could only code for the NS2 protein. Different constructs, leaving Inr1 and its upstream elements intact (for NS1 transcription) but with mutations in the NS2 promoter and NS1/2 translation elements, were made with a luciferase reporter gene (Figure 3-8B and C). Constructs containing the intact luciferase AUG (F1 to F7) served as a positive control for transcription, whereas those lacking the luciferase AUG (F1- to F7-) showed the impact of the various mutations.

P7/7.4 consisted of three segments (Figure 3-8B). The KpnI-SacI segment I contained upstream promoter elements, the SacI-PstI segment II included both TATA boxes, ATG_{NS1}, and both Inrs and, finally, the PstI-HindIII segment III (about 20 nt for the NS1 ORF and about 70 nt for the NS2 ORF) contained the downstream promoter elements (DPE) and ATG_{NS2}. These segments were connected in pBluescript (PCR with primers containing appropriate restriction sites and pCR2.1-AalDNV template) and were transferred as a whole (or after mutation) to the pGL4.20 luciferase reporter. The diagram in Figure 3-8C delineates the different constructs and knockouts (using the Transformer kit according to Clontech's instructions) (Figure 3-8C). Certain ATGs were mutated to TTG, Inr (CATG) was mutated to GCCG, and the TATA box of NS2 was mutated to GCTCGAG. In addition, alternates of F1 to F7 that lacked the luciferase initiation codon (F1- to F7-) as well as two constructs from which the NS2 core promoter was mutated (F2* and F5-*) were obtained. C6/36 and 293T (as for HeLa) cells were transfected, and luciferase activity was determined. Except for F2* and F5-*, all constructs were expected to yield both NS1 and NS2 mRNAs. The observed luciferase activity in C6/36 and 293T cells (Figure 3-8E) corresponded well with the expected luciferase activities, summarized in Figure 3-8D, from (i) fusion proteins with NS1 (F1 and F1-) and NS2 (F5 and F5-), (ii) directly from the luciferase initiation codon (F1 to F7), or (iii) after leaky scanning (for F1, F2, and F5 to F7). Therefore, individual elements of both NS1 and NS2 promoters contributed to mRNA expression.



	F1	F1-	F2	F2-	F3	F3-	F4	F4-	F5	F5-	F6	F6-	F7	F7-
Transcript NS1: ATG (luc)	l.s.	–	l.s.	–	+	–	+	–	l.s.	–	l.s.	–	l.s.	–
fusion NS1	+	+	–	–	–	–	–	–	–	–	–	–	–	–
Transcript NS2: ATG (luc)	+	–	+	–	+	–	+	–	l.s.	–	l.s.	–	l.s.	–
fusion NS2	–	–	–	–	–	–	–	–	+	+	–	–	–	–
Overall expected activity	+	+	+	–	+	–	+	–	+	+	+	–	+	–

Figure 3-8. Analysis of AalDNV P7/7.4 promoter elements using a luciferase reporter gene. (A) Diagram of ORFs in AalDNV, where nt 77 to 423 contain typical sequences of core promoter elements for NS1 and NS2. (B) Three segments (I to III) were connected via pBluescript in pGL4 and could be individually mutated and swapped. (C) In the constructs, filled boxes represent elements thought to direct NS1 mRNA expression and open boxes those thought to direct NS2 mRNA expression, filled triangles represent transcript starts that were established, crosses represent knockouts, and gray boxes represent replaced sequences. Additionally, mutants were made for all constructs in which the initiation codon of luciferase was knocked out (F1- to F7-), and in 2 constructs TATA2 and Inr2 (for NS2 transcripts) were knocked out (indicated by *). (D) Expected activities using the luciferase reporter gene in the pGL4 vector. ("l.s." represents luciferase activity if leaky scanning occurs.) (E) The observed luciferase activity matched the expected activity, except for the very low F7- activity (none expected). These results indicated that the NS initiation codons, and hence the two sets of promoter elements for transcription, were individually important for expression but nevertheless leaky scanning on the NS1 transcript could also contribute to the expression of NS2.

Nevertheless, NS2 could be generated on the NS1 transcript by leaky scanning. Blocking NS2 transcription by mutating its core promoter in F5-* yielded an 90% drop in luciferase activity in C6/36 cells compared to F5-. The remaining activity could be ascribed to a fusion protein from ATG_{NS2} on the NS1 transcript since the introduction of an NS2 stop codon in frame after the NS2 initiation codon in F6- or knocking out ATG_{NS2} in F7- abolished luciferase activity. Consequently, for F2* the remaining luciferase activity was due mainly to leaky scanning on the NS1 transcript and initiation from ATG_{luc}. In conclusion, although NS1 and NS2 mRNAs have their own promoter elements in p7/7.4, NS1 mRNA contributes significantly through leaky scanning to NS2 expression.

In conclusion, Northern blotting, RACE, amplicon mapping, and RPA results were all in agreement. 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}. 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 (16, 17). 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 (18), to regulate NS and VP expression.

Acknowledgments

This work was supported by a grant from the Natural Sciences and Engineering Research Council of Canada to P.T. and scholarships from the Agence Universitaire de la Francophonie and the Fondation Armand-Frappier to H.T.P.

References

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.

2. 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.

3. Bergoin M, Tijssen P. 2010. Densoviruses : a highly diverse group of arthropod parvoviruses, p. 57-90. *In* Asgari S, Johnson KN (ed.), Insect Virology. Horizon Scientific Press, Norwich, U.K.

4. Boublik Y, Jousset FX, Bergoin M. 1994. Complete nucleotide sequence and genomic organization of the *Aedes albopictus* parvovirus (AaPV) pathogenic for *Aedes aegypti* larvae. Virology 200:752-763.

5. Afanasiev BN, Galyov EE, Buchatsky LP, Kozlov YV. 1991. Nucleotide sequence and genomic organization of *Aedes* densonucleosis virus. Virology 185:323-336.

6. Ren X, Hoiczyk E, Rasgon JL. 2008. Viral paratransgenesis in the malaria vector *Anopheles gambiae*. PLoS Pathog 4:e1000135.

7. Jousset FX, Barreau C, Boublik Y, Cornet M. 1993. A parvo-like virus persistently infecting a C6/36 clone of *Aedes albopictus* mosquito cell line and pathogenic for Aedes aegypti larvae. Virus Res 29:99-114.

8. Barreau C, Jousset FX, Bergoin M. 1996. Pathogenicity of the *Aedes albopictus* parvovirus (AaPV), a denso-like virus, for Aedes aegypti mosquitoes. J Invertebr Pathol 68:299-309.

9. Barreau C, Jousset FX, Bergoin M. 1988. Venereal and vertical transmission of the Aedes albopictus parvovirus in Aedes aegypti mosquitoes. Am J Trop Med Hyg 57:4364-4370.

10. Ward TW, Kimmick MW, Afanasiev BN, Carlson JO. 2001. Characterization of the structural gene promoter of *Aedes aegypti* densovirus. J Virol 75:1325-1331.

11. Kimmick MW, Afanasiev BN, Beaty BJ, Carlson JO. 1998. Gene expression and regulation from the p7 promoter of *Aedes* densonucleosis virus. J Virol 72:4364-4370.

12. Afanasiev BN, Kozlov YV, Carlson JO, Beaty BJ. 1994. Densovirus of *Aedes aegypti* as an expression vector in mosquito cells. Exp Parasitol 79:322-339.

13. Tijssen P, Li Y, El-Far M, Szelei J, Letarte M, Zadori Z. 2003. Organization and expression strategy of the ambisense genome of densonucleosis virus of *Galleria mellonella*. J Virol 77:10357-10365.

14. 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.

15. Fediere G, El-Far M, Li Y, Bergoin M, Tijssen P. 2004. Expression strategy of densonucleosis virus from *Mythimna loreyi*. Virology 320:181-189.

16. Kutach AK, Kadonaga JT. 2000. The downstream promoter element DPE appears to be as widely used as the TATA box in Drosophila core promoters. Mol Cell Biol 20:4754-4764.

17. Burke TW, Willy PJ, Kutach AK, Butler JE, Kadonaga JT. 1998. The DPE, a conserved downstream core promoter element that is functionally analogous to the TATA box. Cold Spring Harbor symposia on quantitative biology 63:75-82.

18. Hsu JY, Juven-Gershon T, Marr MT, 2nd, Wright KJ, Tjian R, Kadonaga JT. 2008. TBP, Mot1, and NC2 establish a regulatory circuit that controls DPE-dependent versus TATA-dependent transcription. Genes & development 22:2353-2358.

3.1.5. Consequences of findings on PstDNV classification

Based on the genome sizes and open reading frame structure, the shrimp PstDNV and mosquito AalDNV were tentatively classified into the same genus *Brevidensovirus*. Few details with respect to the expression of these parvoviral genomes are documented. Here we report the complete transcription maps of PstDNV and AalDNV by Northern blotting, RACE analysis, RNase protection assays, amplicon mapping and sequencing of their transcripts. The transcription of PstDNV was found to be under the control of tree distant promoters P2, P12 and P61 resulting in mRNAs for NS1 (3362 nts), for NS2 (3131 nts) and VP (1149 nts), respectively. AalDNV also produced three mRNAs of similar sizes (3359, 3345, and 1246 nts) but by a different expression strategy. The AalDNV transcription strategy differed from that of PstDNV by not using splicing for the NS1 transcripts. In contrast to widely spaced NS promoters of PstDNV, the overlapping P7/7.4 promoters of AalDNV employed closely located alternate transcription initiation sites, positioned at either side of the NS1 initiation codon. All NS mRNAs co-terminated with the VP mRNA. Activities of the promoters of these two densoviruses were explored using luciferase assays and were found functional in both insect and human cell lines.

Although the genome size and ORFs of PstDNV resembled that of AalDNV and other brevidensoviruses, the expression strategy is different and indicates that PstDNV should not be classified as a brevidensovirus. NS1 of PstDNV and AalDNV, despite a low sequence identity, contained replication initiation motifs and DNA-dependent ATPase motifs (helicase SF3) (Fediere *et al.*, 2002) but these motifs are not restricted to parvoviruses and are found in diverse replicons from eukaryotes to eubacteria and a wide range of viruses (Gorbalenya *et al.*, 1989, Ilyina *et al.*, 1992). These two viruses do not contain the PLA2 motif found in the VP of most parvoviruses (Canaan *et al.*, 2004, Zadori *et al.*, 2001).

Moreover, so far terminal hairpins have not been found in the PstDNV genome although there are long direct terminal repeats of 81 nts, suggesting that PstDNV may use DTRs for rolling-circle replication instead of hairpins for other parvoviruses. Further research is necessary in order to confirm this hypothesis.

The significant differences found in transcription and probably telomere sequences of AalDNV and PstDNV, despite their similarities in their genome organization, indicated that PstDNV should be classified as a new genus of densovirus.

3.2. Section 2: Expression strategy of Pefudensovirus, *Acheta domesticus* densovirus (AdDNV), and genomic comparison of AdDNV isolates related to different outbreaks in Europe, North America and Japan

3.2.1. Résumé en français des différents travaux sur la stratégie d'expression du Pefudensovirus du grillon *Acheta domesticus* (AdDNV) et la comparaison des isolats d'AdDNV responsables de différentes épidémies survenues en Europe, en Amérique du Nord et au Japon.

Introduction

L'AdDNV isolé en France en 1977 est endémique en Europe depuis au moins 36 ans où il a été responsable de plusieurs épizooties dans les élevages de masse, notamment en Angleterre et en Allemagne. Nous rapportons, dans la Publication 3, les résultats du clonage, du séquençage et de l'analyse des transcrits de la souche originelle de l'AdDNV et démontrons que ce virus a développé une stratégie d'expression unique parmi les densovirus. L'élevage industriel du grillon domestique servant au nourrissage d'animaux de compagnie ou comme appâts représente un chiffre d'affaires de plusieurs millions de dollars par année en Amérique du Nord. Les épizooties catastrophiques survenues dans les années 2009-2010 dans ces élevages ont conduit à l'effondrement de ce commerce et à l'importation de nouvelles espèces de remplacement. Enfin, les résultats de l'analyse génomique comparée de 8 isolats d'AdDNV provenant d'épizooties survenues en Europe, au Japon et en Amérique du Nord sont présentés dans la Publication 4.

Publication 3: Liu K, Li Y, Jousset FX, Zadori Z, Szelei J, Yu Q, Pham HT, Lépine F, Bergoin M, Tijssen P., J Virol. 2011 Oct; 85(19):10069-78. **Le densovirus d'*Acheta domesticus* isolé du grillon domestique d'Europe a développé une stratégie d'expression unique parmi les densovirus.**

Depuis 2009, l'AdDNV a causé d'importants dégâts dans les élevages industriels de cet insecte aux États-Unis et au Canada. Le génome de la souche européenne de l'AdDNV a été cloné et séquencé. Sa taille de 5425 nucléotides inclut une répétition inversée de 144 nts à chaque extrémité dont les 114 premiers peuvent se replier pour former une structure de type épingle à cheveux. L'organisation des gènes est de type ambisens, avec trois cadres de lecture ouverts (ORFs) dans la moitié 5' d'un brin, codant pour les protéines non-structurales NS1, NS2 (dont le cadre de lecture est inclus dans celui de NS1) et NS3. Deux ORFs, l'un de grande taille

(A) et l'autre de petite taille (B) occupent la moitié 5' du brin complémentaire et codent pour les protéines de capside (VP). Ce type d'organisation des gènes est typique des densovirus appartenant au genre *Pefudensovirus.* Les ARNs messagers viraux ont été identifiés à partir d'ARN total extrait de tissus adipeux de grillons infectés et leur cartographie établie par les techniques de Northern blot et RACE. Les trois protéines NS sont exprimées par deux transcrits; l'un non épissé de 2.5 kb servant à exprimer la protéine NS3, l'autre de 1,8 kb résultant d'un épissage du transcrit de 2.5 kb éliminant la séquence codante de NS3. Ce transcrit sert à produire la protéine NS1 et la protéine NS2 par «leaky scanning», l'environnement peu favorable de l'AUG de NS1 et celui beaucoup plus favorable de l'AUG de NS2 favorisent ce type d'expression. Ces deux transcrits sont sous contrôle d'un promoteur localisé juste en aval de la fin de l'ITR et co-terminent en aval de la séquence codante de NS1. Ce mode d'expression des gènes NS de l'AdDNV est semblable à celui décrit chez les DNVs du genre *Densovirus*. L'élucidation des modalités d'expression des protéines VP s'est avérée plus délicate. En effet, l'analyse par Northern blot à l'aide d'une sonde VP ne détecte qu'un seul messager d'environ 2.5 kb. Or, l'identification d'une protéine de capside de 110 kDa d'une part et la détection dans l'ORF B (en position 5'-terminale) de la séquence PLA2 caractéristique de la région N-terminale de la protéine VP1, présente chez tous les densovirus à génome ambisens d'autre part impliquait un épissage permettant de mettre en phase les ORFs B et A. La séquence de la protéine VP2 identifiée par dégradation N-terminale selon Edman et par la spectrométrie de masse, correspond à l'ORF A. Dans un premier temps, a recherche d'épissage à l'aide d'un couple d'amorces prises dans la région 5' terminale de l'ORF A (en aval de l'AUG de VP2 et 5' terminale de l'ORF B (en aval de l'AUG de VP1) a permis de détecter deux types d'épissages alternatifs Ia et Ib dans la séquence 5' terminale de l'ORF B à partir de deux sites donneurs et d'un site accepteur tous situés en amont de l'AUG de VP2. Toutefois, aucun de ces deux épissages ne mettait en phase les cadres de lecture B et A. L'expression dans un système baculovirus de la région supposée de l'épissage mettant en phase les deux ORFs et l'analyse par spectrométrie de masse et par Western blot des protéines exprimées ont permis de démontrer que le nucléotide 4435 localisé dans l'ORF B était connecté en phase avec le nucléotide 4259 localisé dans l'ORF A. Cet épissage (II), qui a été par la suite confirmé par RT-PCR à l'aide d'amorces appropriées, met en phase la fin de la séquence de l'ORF B avec la région N-terminale de l'ORF A en éliminant l'AUG d'initiation de VP2. Le site donneur de cet épissage étant situé dans la séquence commune des deux introns (Ia et Ib), les épissages Ia-Ib et II sont exclusifs l'un de l'autre. Les transcrits de type Ia-Ib servent donc à exprimer VP2 et vraisemblablement VP3 et VP4 par leaky scanning alors que les transcrits de type II servent

à exprimer VP1, et probablement VP3 et VP4, également par leaky scanning. Ce type d'expression des protéines de structure de l'AdDNV est unique chez les parvovirus et définit un nouveau genre de densovirus à génome ambisens.

Publication 4: Pham HT, Iwao H, Szelei J, Li Y, Liu K, Bergoin M, Tijssen P.,Genome Announc. 2013 Aug 15 ; 1(4). **L'analyse génomique comparée d'isolats du densovirus d'*Acheta domesticus* provenant de différentes épizooties survenues en Europe, en Amérique du Nord et au Japon.**

Nous rapportons dans cette publication les résultats du clonage et du séquençage complet des génomes de 8 isolats d'AdDNV responsables d'épizooties dans différentes régions du globe que nous comparons à celui du premier isolat de 1977. Il s'agit de 4 épizooties survenues en juin 2004, juillet 2006, mai 2007 et août 2009 dans un élevage *d'A. domesticus* en Allemagne, deux épizooties survenues en Amérique du nord, l'une en septembre 2012 au Canada dans un élevage de *Gryllus sigillatus*, l'autre aux États-Unis en mars 2012 dans un élevage de *Gryllus locorojo* et enfin une épizootie survenue au Japon en décembre 2012 dans un élevage d'*A. domesticus*. Tous les isolats présentent la même taille de 5425 nucléotides et la même longueur de 144 nt pour leur ITR. L'organisation de leurs séquences codantes des promoteurs et des sites d'épissage sont identiques à ceux de l'isolat de 1977. Les plus fortes identités concernent les séquences des protéines NS1 et NS2 (99.3%), tandis que les séquences de la protéine NS3 présentent 94% d'identités et celles des protéines de capside 98.1% (ORF A) et 97.1% (ORF B) d'identités. L'analyse phylogénétique des sequences complètes des 8 isolats montre que les 4 souches européennes et la souche américaine sont très voisines mais que la souche japonaise et la souche canadienne ont divergé précocement des souches européennes et américaines; au moins 20 ans avant que ne se produisent simultanément les épizooties de 2009 en Amérique du Nord et au Japon. Ces deux épizooties semblent donc sans lien l'une avec l'autre.

3.2.2. Contribution of authors to publications in Section 2

Publication 3: 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. Zadori Z and Szelei J did RACE and promoter activity experiments. Qian Yu contributed to make and express VP clones, purified native viruses. Pham HT carried out SDS-PAGE, Western blot and prepared samples for mass spectrometry. Lépine F was responsible

for mass spectrometry analysis. Peter Tijssen and Max Bergoin planned and supervised the project. Peter Tijssen prepared the manusript. Other authors contributed to review the manuscript before submission.

Publication 4: Iwao H provided infected crickets from Japan. Szelei, Li Y and Liu K did initial PCR and sequencing on five AdDNV isolates from *A. domesticus* of different outbreaks in Europe and North America. Pham HT was responsible for other three isolates from Japanese *A. domesticus,* Canadian *Grylodes sigillatus,* and American *Gryllus locorojo*. Peter Tijssen supervised the project. Pham and Tijssen prepared the manuscript.

3.2.3. Publication 3: 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 FX[3], Zadori Z[1], Szelei J[1], Yu Q[1], Pham HT[1], Lépine F[1], Bergoin M[1,3], Tijssen P[1]

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

[2]Central China Normal University, Wuhan 430079, People's Republic of China

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

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.

Introduction

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, *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 (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 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 (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 *Pefudensovirus* 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 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 (24).

Virus and DNA preparation. Virus was purified as previously described (29). Lysis buffer [300 l 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 (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-ul 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 BglII-SalI 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 3-3) 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 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 ProVP1F and ProVP1R primers were used for the VP1 promoter, and the PrNSMf and PrNSMr primers were used for the *Mythimna loreyi* densovirus (MlDNV) NS promoter (control) (Table 3-3). 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 5X10^5 cells/ml. The cells were cultured overnight. Transfection was performed with 2.5 μl DOTAP reagent and 0.6 g 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 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 3-3), 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 3-3, 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 Ad1HAR, containing an XbaI site, was used (Table 3-3). 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 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 (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 ingel 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 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 Annexe 2). 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 (Figure 3-9A, 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 3-9A). 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 denso-virus 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 3-9A; see Figure S1 in the Annexe 2). 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 containing the ORFs for the NS genes as the plus strand so the genes would be located similarly.

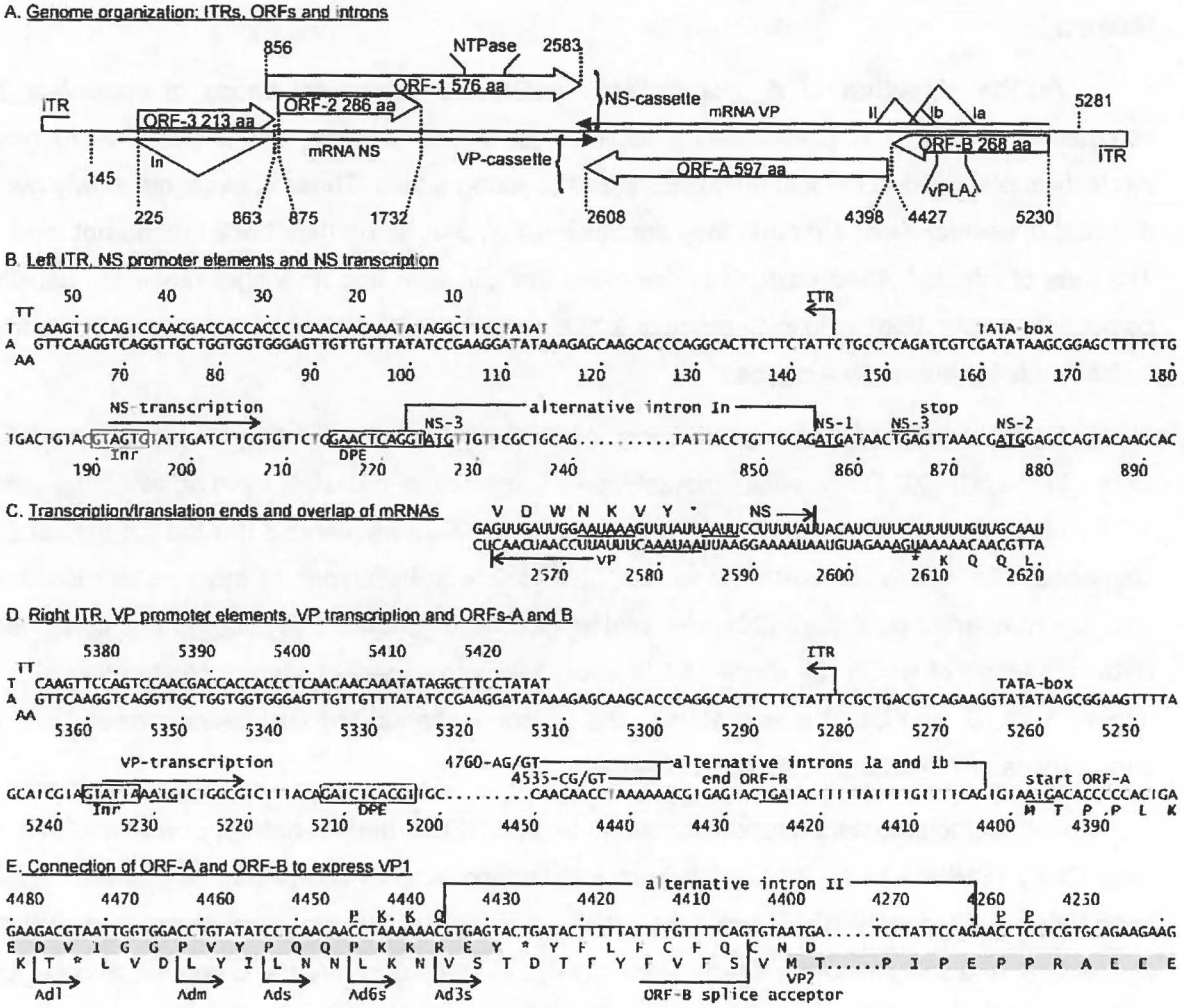


Figure 3-9. Genome organization of AdDNV. (A) Overview of genome organization, positions, and sizes of ITRs, ORFs, and introns. The In intron in the NS mRNA, between nt 223 and nt 855, occurs in about half of the NS transcripts. Ia (nt 4403 to 4758), Ib (nt 4403 to 4533), and II (nt 4260 to 4434) are 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 *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 3-10A), 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 3-9D; see Figure S1 in the Annexe 2), 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 estimated by SDS-PAGE (Figure 3-10A). 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 3-9E; see Figure S1 in the Annexe 2), an unspliced transcript could also code for VP1 by a ribosome frameshift in the ORF-A–ORF-B overlap-ping 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) (Figure 3-10B). 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 3-9B to D. The 3' termini of the NS and VP transcripts had a 34-nt overlap (Figure 3-9C), 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.

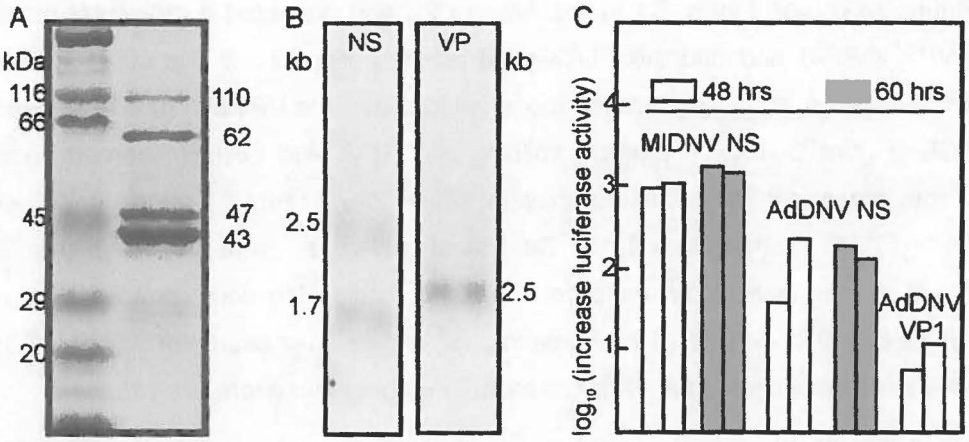


Figure 3-10. (A) SDS-PAGE analysis of structural proteins of AdDNV. Lane 1, Bio-Rad broad-range standard proteins as markers; lane 2, four proteins of AdDNV (VP1 to -4, in decreasing size). The estimated masses corresponded reasonably with the sequence-predicted masses (88, 65, 51, and 47 kDa, respectively). (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.

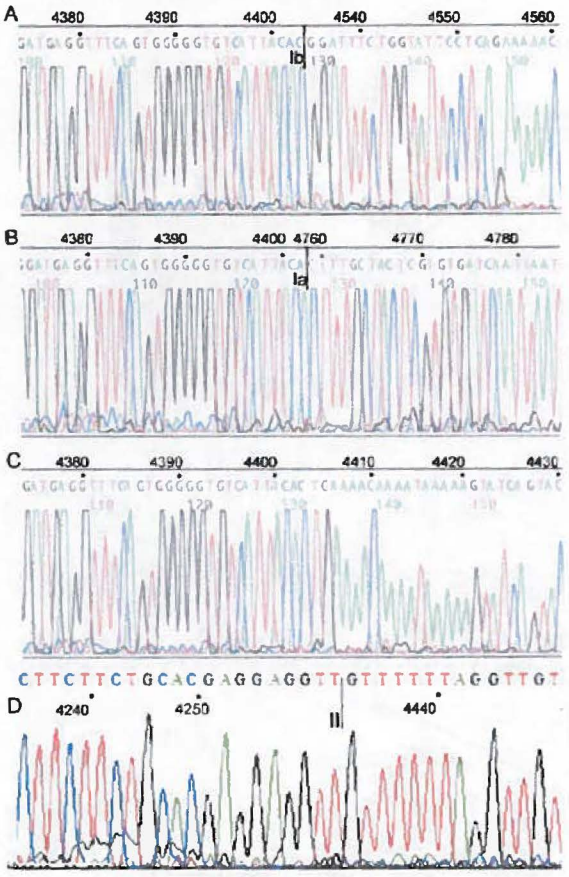
Table 3-3. PCR primers used in this study

Primer	Sequence[a]	Position (nt) in AdDNV	Target or use
AdVPR	TTTGTTGCAATCCCATAATAGTAC	2610–2633	Near 3 end of VP mRNA
NAdR	gctctagatCATCTTGAACGTTTACCACCACT	3892–3915	Just upstream of VP4
Adsp	tcggaattcCACGTTCTTGTGGATGAGG	4362–4380	19-37 nt into VP2
Adsvp	gccTACCAGAAATCCGTGTAATGACA	4546–4534/4403–4393	Small intron splice
Ad3s	CGTGAGTACTGATACTTTTTATTT	4435–4412	End of ORF-B (TGA)
Ads	gccCCTCAACAACCTAAAAAACGTGAGTACTGA	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	gccCCTGTATATCCTCAACAACCTAAA	4462–4439	End of ORF-B (4 aa)
AdlgF	tcggaattcATGTCTGGCGTCTTTACA	5230–5213	Start of VP1 (ORF-B)
RNAtag	gggtctagagctcgagT_{17}	Poly(A) sequence	3 /5 -RACE (first round)
ADAP	gggtctagagctcgagT		Subsequent rounds of RACE
ProNSF	acggtaccGATATAAAGAGCAAGCACCC	109–128	NS1 promoter (KpnI)
ProNSR	gaagatCTGTGCTGGAGGCGCTTCTACTGCAGCGAAAAC GTACCTGAGTTCCAGAACAC	240–207	NS1 promoter (reverse)
ProVP1R	gaagatctAAGACGCCAGAGATTTAATACT	5217–5238	VP1 (BglII site; reverse)
ProVP1F	gtaggtaccGATATAAAGAGCAAGCACCCA	5323–5300	VP1(KpnI site; forward)
PrNSMf	ACGGTACCGACTATAAATAGAGCTGAGC		MIDNV (forward)
PrNSMr	GAAGATCTATCTTGCAATAGATATACCTA		MIDNV (reverse)
pFECRIF	CGCAAATAAATAAGAATTCTACTGTTTTCGTAAC		Mutation in pFastBac1
pFECRIR	GTTACGAAAACAGTAGAATTCTTATTTATTTGCG		Mutation in pFastBac1
Ad1HAR	GCTCTAGATCAAGCGTAATCTGGAACATCGTATGGGTA TTTTTGTTGCAATCCCATAATA		VP products (reverse primer HA)
AdATG1B	ggaattcATGTCCGTCTTTACAGATCTCAC		
AdmATG1B	ggaattcACCTCCGTCTTTACAGATCTCTCAC		

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

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 3-3). 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 3-9D and 3-11A 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 3-9D). 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 3-3) 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 (Figure 3-9E).

Expression of VP1. Figure 3-9E 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 (Figure 3-12A). The PCR products obtained using the forward primers Ad3s, Ad6s, Ads, Adm, and Adl (Table 3-3) (the distance from the ORF-B stop codon is indicated, in codons, in Figure 3-12A) and the reverse primer NAdR, chosen at the beginning of the extension of VP4 to ensure stable products, 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.



E Mass spectrometry of capsid proteins

VP1

```
  1 MSGVPTDLTL RDFSLGTNNL PLANTPKLRN RFGRWWNRSH FYDRLPTNRP EPLRETSFQE
 61 AAGPESTRID IARDEINAGE GAAQAETSFS TGVESETALEL GEAATETTGL LGGATAGSAA
121 AGTAGTLGTV AATAAGGAAL AGIGIGIFKL IDHTSSKGAV LPGTDFVGPG NPIDPKPARS
181 ETDQIAKEHD LGYEDLLERA KSQTFTEEDF RTEVYKLDDE AIHRFSEEYQ KSGTWQAFVG
241 KYGLKAKALI EDVIGGPVYP QQFKKQAPRA EEEEVAPATP TADDILDELP EGNPEDIDDF
301 DVDQLPSTTQ EHHPTTDLAA QPGRSTDPDF SPVSNPIDST PPDEGDSSSM SAPLTPPITP
361 DKTQDKGKGK RGGGGRPPKS SGGKRSRMAG TSLPGAGADI AGRGGGQAII PIPRPITTPH
421 KNFIIFNKVH RFLTTGLANV VIPITRTVNE VTTVDNFVIT SLGRLPVDRP FLYNNPNEFA
481 QLPPGSCIEE CNVRVTAFTP RIAFQTNSSH TGLATLNQMQ FILEATGLNI KTQGVDVRPK
541 EFQANEPMVV SSIDELGAIR DENLFEDIVK EFYGDKSVPN HVPRHQFGIP YPLKNYPALV
601 LQQNGTADNP GYECLQSHVN EIRCNGPPGE IVEYNYKPQL GLIKKAIPAV YTGVPSAVGG
661 NSTLSCPAGA GNSQYRTADF TMTNNILRGD SETFKDTVVS NTEWTIATQI EKSQILHAGA
721 FPSITPRAQP SLHIGVKPVH ALTTANLDDN LNNANFIDTQ AYFTVKAACK VRLQYPSIRP
781 LFNGPNTIPD NQIYSSSTLQ YTHAVGKSTI MGLQQK
```

VP2

```
  1 MTPPLKPHPQ ERDNNEYLNE GQRRIAVEQW QLARVRRGLP IDHPIPEPPR AEEEEVAPAT
 61 PTADDILDEL PEGNPEDIDD FDVDQLPSTT QEHHPTTDLA AQPGRSTDPD PEPVSNPIDS
121 TPPDEGDSSS MSAPLTPPIT PDKTQDKGKG KRGGGGRPPK SSGGKRSRMA GTSLPGAGAD
181 IAGEGGGQAI IPIPRPITTP HNNFIIFNKV HRFLTTGLAN VVIPITRTVN DVTTVDNFVI
241 TSLGRLPVDR PFLYNNPNEF AQLPPGSCIE ECNVRVTAFT PRIAFQTNSS NTGLATLNQM
301 QFILEATGLN IKTQGVDVRP KEFQANEPMV VSSIDELGAI RDENLFEDIV KEFYGDKSVP
361 NHVPRHQFGI PYPLKNYFAL VLQQNGTADN PGYECLQSHV NEIRCNGPPG EIVEYNYKPQ
421 LGLIKKAIPA VYTGVPSAVG GNSTLSCPAG AGNSQYRTAD FTMTNNILRG DSETFKDTVV
481 SNTEWTIATQ IEKSQILHAG AFPSITPRAQ PSLHIGVKPV HALTTANLDD NLNNANFTDT
541 QAYFIVEAAC KVRLQYPSIR PLFNGPNTIP DNQIYSSSTL QYTHAVGKST IMGLQQK
```

VP3

```
  1 MSAPLTPPIT PDKTQDKGKG KRGGGGRPPK SSGGKRSRMA GTSLPGAGAD IAGEGGGQAI
 61 IPIPRPITTP HNNFIIFNKV HRFLTTGLAN VVIPITRTVN DVTTVDNFVI TSLGRLPVDR
121 PFLYNNPNEF AQLPPGSCIE ECNVRVTAFT PRIAFQTNSS NTGLATLNQM QFILEATGLN
181 IKTQGVDVRP KEFQANEPMV VSSIDELGAI RDENLFEDIV KEFYGDKSVP NHVPRHQFGI
241 PYPLKNYFAL VLQQNGTADN PGYECLQSHV NEIRCNGPPG EIVEYNYKPQ LGLIKKAIPA
301 VYTGVPSAVG GNSTLSCPAG AGNSQYRTAD FTMTNNILRG DSETFKDTVV SNTEWTIATQ
361 IEKSQILHAG AFPSITPRAQ PSLHIGVKPV HALTTANLDD NLNNANFTDT QAYFIVEAAC
421 KVRLQYPSIR PLFNGPNTIP DNQIYSSSTL QYTHAVGKST IMGLQQK
```

VP4

```
  1 MAGTSLPGAG ADIAGKGGGQ AIIPIPRPIT TPHNNFIIFN KVERFLTTGL AMVVIPITRT
 61 VNDVTTVDNF VITSLGRLPV DRPFLYNNPN EFAQLPPGSC IEECNVRVTA FTPRIAFQTN
121 SSNTGLATLN QNQFILHATG LNIKTQGVDV RPKEFQANEP MVVSSIDELG AIRDENLFED
181 TVKEPYGDKS VPNHVPRHQF GIPYPLKNYF ALVLQQNGTA DNPGYECLQS HVNEIRCNGP
241 PGEIVEYNYK PQLGLIKKAI PAVYTGVPSA VGGNSTLSCP AGAGNSQYRT ADFTMTNNIL
301 RGDSETFKDT VVSNTEWTIA TQIEKSQILH AGAFPSITPR AQPSLHIGVK PVHALTTANL
361 DDNLNNANFT DTQAYFIVEA ACKVRLQYPS IRPLFNGPNT IPDNQIYSSS TLQYTHAVGK
421 STIMGLQQK
```

Figure 3-11. (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.

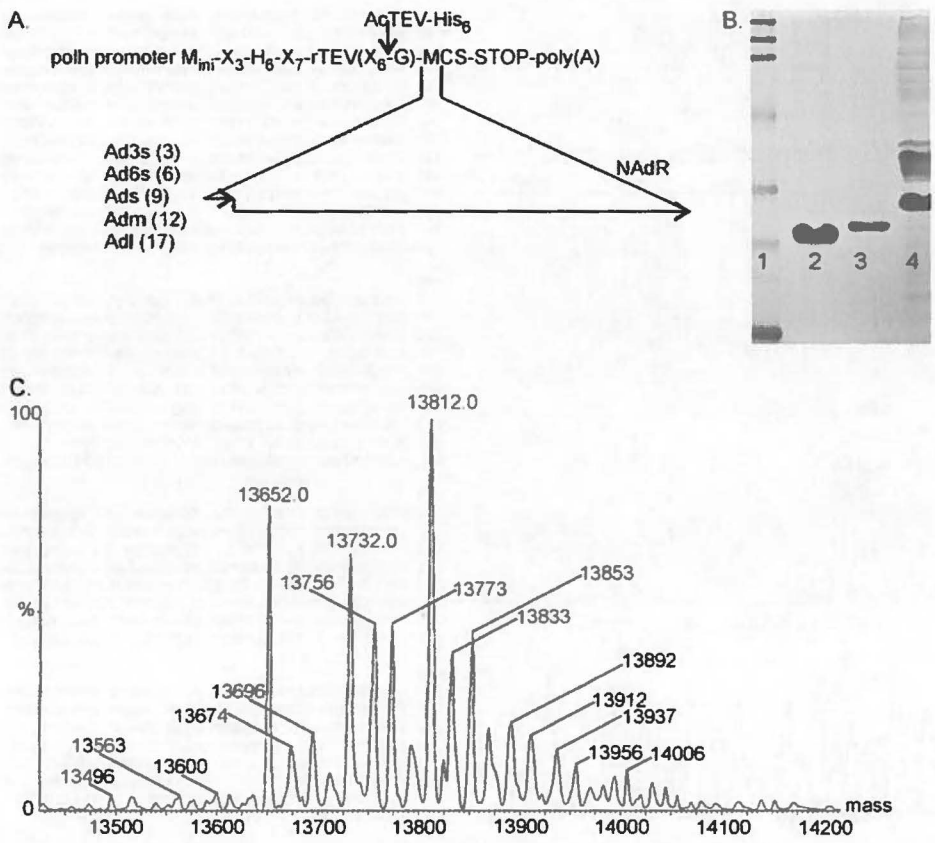


Figure 3-12. (A) Construction of various ORF-A and ORF-B constructs with different Ad primers (Ad3s, Ad6s, etc.) linking a short part of ORF-B to a larger part of ORF-A, using the NAdR primer and the pFastbacHTb baculovirus system. Upstream of the insert is a TEV proteolytic site and a histidine tail (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-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 (Figure 3-12B). 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 3-12C). 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 3-10A), 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 3-3), 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 3-11D). As illustrated in Figure 3-9D 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 3-13A). 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 3-13A), 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 3-13B).

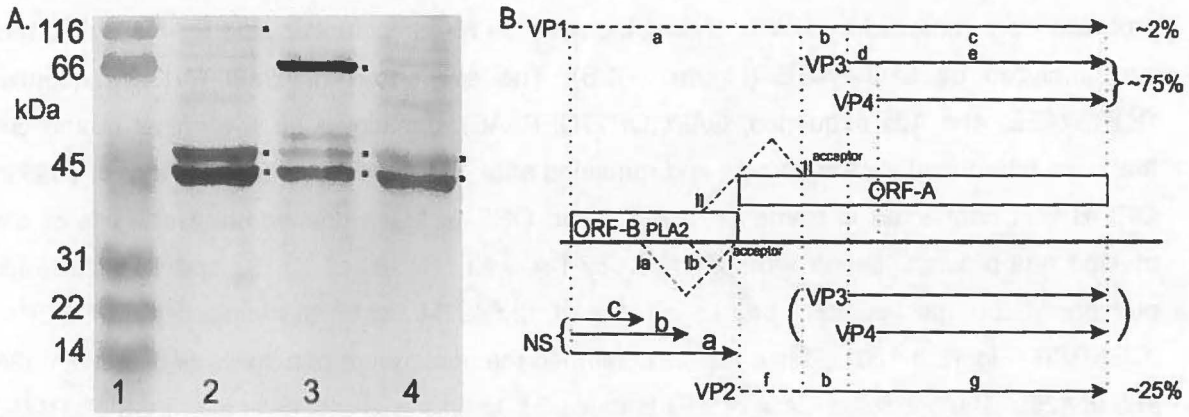


Figure 3-13. (A) Western blot of pFastbac VP constructs. Lane 1, prestained SDS-PAGE protein standards (Bio-Rad). Lane 2, expression products from the ORF-A–ORF-B construct from which intron II was deleted. Strong bands were observed for VP3 and VP4, with a weak band for VP1 (all indicated with dots), 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 Figure 3-10A). 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 3-9B 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 3-10C). 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 3-11E). 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 3-14), confirming that the 816-aa species corresponded to VP1.

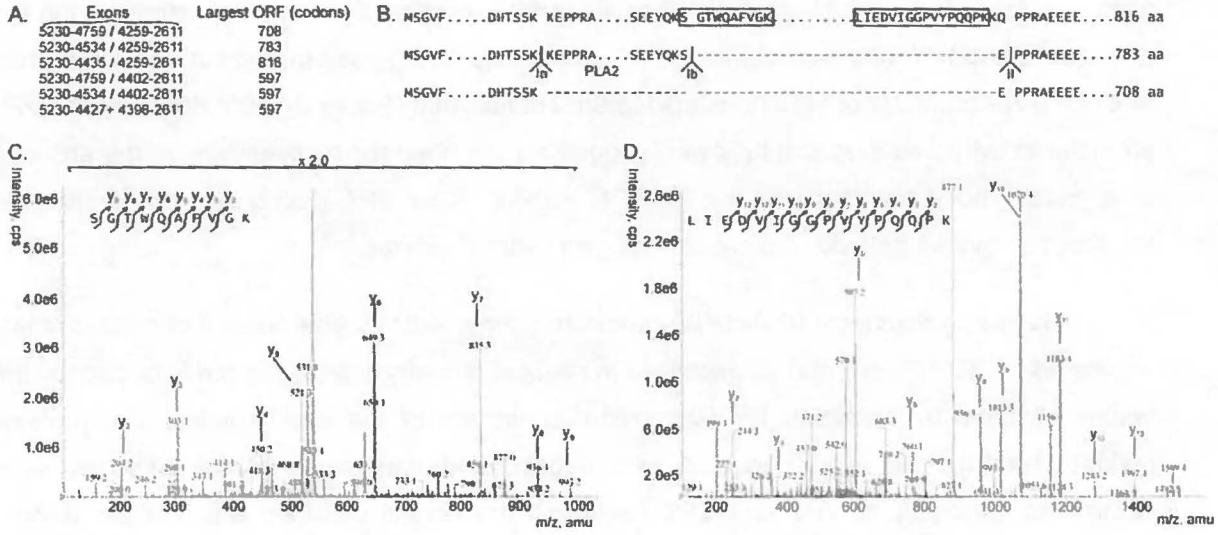


Figure 3-14. Potential VP1 proteins encoded by the VP gene cassette. (A) Different potential splice combinations of the ORFs in the VP gene cassette and the largest protein that could then be coded. N-terminal amino acid sequencing showed that the 597-codon ORF corresponded to VP2 and, theoretically, 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 SGTWQAFVGK 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 [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 Figure S3 in the supplemental material, Annexe 2). 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 3-10C). 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 3-9A), 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 (Figure 3-11C), yielding a 268-aa ORF-B product (NSa) that is not involved in capsid formation (Figure 3-9A). 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 3-14).

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

1. 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.

2. 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.

3. 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.

4. 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.

5. Fediere, 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.

6. Fediere, G., Y. Li, Z. Zadori, J. Szelei, and P. Tijssen. 2002. Genome organization of *Casphalia extranea* densovirus, a new iteravirus. Virology 292:299–308.

7. Fernandes, S., and P. Tijssen. 2009. Seamless cloning and domain swapping of synthetic and complex DNA. Anal. Biochem. 385:171–173.

8. 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.

9. 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.

10. Kaufmann, B., et al. 2010. Structure of *Penaeus stylirostris* densovirus, a shrimp pathogen. J. Virol. 84:11289–11296.

11. Kaufmann, B., et al. 2011. Structure of *Bombyx mori* densovirus 1, a silkworm pathogen. J. Virol. 85:4691–4697.

12. 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.

13. Laemmli, U. K. 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature 227:680–685.

14. Luckow, V. A., S. C. Lee, G. F. Barry, and P. O. Olins. 1993. Efficient generation of infectious recombinant baculoviruses by site-specific trans-poson-mediated insertion of foreign genes into a baculovirus genome propagated in Escherichia coli. J. Virol. 67:4566–4579.

15. Matsudaira, P. 1987. Sequence from picomole quantities of proteins electroblotted onto polyvinylidene difluoride membranes. J. Biol. Chem. 262: 10035–10038.

16. Meynardier, G., et al. 1977. Virus de type densonucléose chez les orthoptères. Ann. Soc. Entomol. Fr. 13:487–493.

17. 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 *Blattella germanica*: detection, nucleotide sequence and genome organiza-tion. 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. Ross-mann. 1998. The structure of an insect parvovirus (*Galleria mellonella* densovirus) at 3.7 A 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. Com-plete nucleotide sequence and genomic organization of hepatopancreatic parvovirus (HPV) of *Penaeus monodon*. Virology 346:266–277.

24. Szelei, J., et al. 2011. Susceptibility of North-American and European crick-ets 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 (paravovirus). 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.

3.2.4. Publication 4: Comparative genomic analysis of *Acheta domesticus* densovirus isolates from different outbreaks in Europe, North America, and Japan

Pham HT[1], Iwao H[2], Szelei J[1], Li Y[3], Liu K[4], Bergoin M[1], Tijssen P[1]

Genome Announc. 2013 Aug 15; 1(4)

[1]INRS, Institut Armand-Frappier, Laval, QC, Canada

[2]Marinpia Nihonkai 5932-445, Nishifunami, Niigata, Japan

[3]Bioengineering Department, Wuhan Bioengineering Institute, Yangluo, Wuhan, Hubei, People's Republic of China

[4]College of Life Sciences, Central China Normal University, Wuhan, Hubei, People's Republic of China

Abstract

The first densovirus from a cricket, *Acheta domesticus* densovirus (AdDNV) (*Parvoviridae*), was isolated in Europe in 1977 and has been studied previously. We compared seven additional AdDNV genomes isolated from 4 other European outbreaks, 2 major North American outbreaks, and a Japanese outbreak. Phylogenetic analysis suggested that the 2009 Japanese and North American outbreaks were not related.

The United States cricket pet food industry supplies billions of live crickets (*Acheta domesticus* L.) per year to pet stores and individuals. This industry has been devastated by epizootic *Acheta domesticus* densovirus (AdDNV) outbreaks since 2009 (1). Alternative, presumably virus-resistant, field cricket species have been introduced and led to widespread United States (and European) distribution of exotic *Gryllus* species, such as the naturally widespread African, European, and Asian "black cricket," *G. bimaculatus*, and the previously unknown red cricket, *G. locorojo*, despite existing United States federal regulations to prevent such movement (2).

Previously, AdDNV outbreaks have occurred in Europe, and a small, contained one occurred in an operation in the Southeastern United States (3). The isolate from this U.S. outbreak, preserved in paraffin, was sequenced and was found to be identical to the European isolate from 1977 (1). The 1977 isolate (4) was cloned and sequenced and its expression strategy studied (5). In particular, expression of the structural proteins was found to differ from those of other parvoviruses, since two large open reading frames (ORFs) were spliced and VP2 was not completely contained within VP1, i.e., structural proteins did not form a nested set of N-extended proteins like those of other densoviruses. The *Blattella germanica* densovirus (BgDNV) was found to have adapted a similar strategy (6).

AdDNV samples were obtained from different outbreaks: June 2004 (AdEu04), July 2006 (AdEu06), May 2007 (AdEu07), and August 2009 (AdEu09), all in Germany; September 2009 (AdNA09) in the United States and Canada; and December 2012/ January 2013 in Japan (AdJP12). Moreover, AdDNV was isolated from different cricket species, *Gryllodes sigillatus* in September 2012 in Canada (GsCa12) and *Gryllus locorojo* in the United States in March 2012 (GlUS12). For all isolates, DNA samples were isolated as previously described (5). Primers (5) were designed on the sequence obtained previously for the 1977 AdDNV isolate from Europe (GenBank accession number HQ827781), so that overlapping amplicons could be obtained, whereas termini were cloned as described previously (7, 8).

At least two complete clones of each isolate were sequenced in both directions, and four times at locations of discrepancies, using Sanger's method and the primer-walking method as described previously (9). Contigs were assembled by use of the CAP3 program (http://pbil.univ-lyon1.fr/cap3.php/) (10). Like some other densoviruses (11–16), AdDNV was found to have a broad host range, infecting at least *Acheta domesticus*, *Gryllus locorojo*, and *Gryllodes sigillatus*. In fact, AdNA09 and the GlUS12 isolates had the same sequence.

All isolates had genomes of 5425 nucleotides (nt) and 144 nt-inverted terminal repeats, of which the 114 distal nt formed perfect hairpins. The location of open reading frames, TATA boxes, and splicing sites were all conserved compared to those of the original 1977 isolate (5). The highest protein sequence identity among the isolates was found for NS1 and its overlapping NS2 (both 99.3%), whereas NS3 had an identity of 94.4%, and the two ORFs, A and B, of the structural proteins had identities of 98.1 and 97.1%, respectively. Interestingly, phylogenetic analysis showed that the AdJP12 and GsCa12 clade diverged early from the European/North America clade, probably 20 years before the epidem-ics occurred simultaneously in 2009 in North America and Japan. This suggested another contributing factor to these outbreaks.

Nucleotide sequence accession numbers. The GenBank accession numbers are KF015274 for the AdDNV-AdEu04, KF015275 for the AdDNV-AdEu06, KF015276 for the AdDNV-AdEu07, KF015277 for the AdDNV-AdEu09, KF015278 for the AdDNV-AdNA09 and AdDNV-GlUS12, KF015279 for the AdDNV-AdJP12, and KF015280 for the AdDNV-GsCa12 isolates.

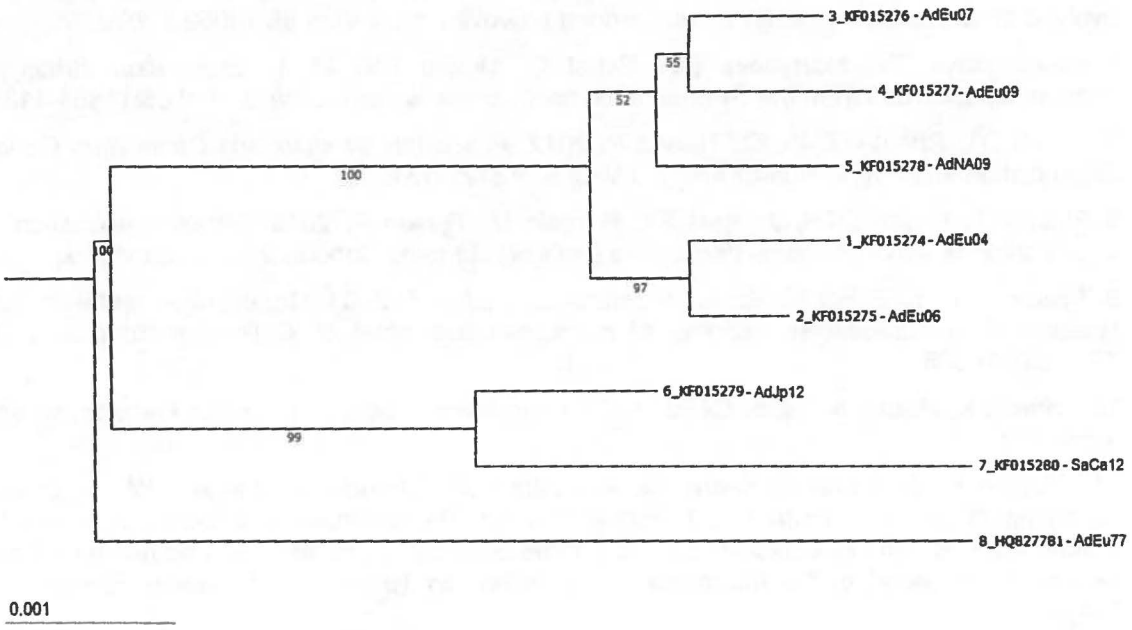


Figure 3-15. Phylogenetic analysis of AdDNV Isolates. The tree was generated from 5425 nts by MAFFT multiple alignment program for nucleotide sequences. Bootstrap values showed on the branches were calculated from 1000 replicates.

Acknowledgments

This work was supported by the Natural Sciences and Engineering Research Council of Canada to P.T.; H.T.P. acknowledges tuition fee waivers at INRS.

References

1. 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.

2. 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.

3. Styer EL, Hamm JJ. 1991. Report of a densovirus in a commercial cricket operation in the Southeastern United-States. J Invertebr Pathol 58:283-285.

4. Meynadier G, Matz G, Veyrunes J-C, Bres N. 1977. Virus de type densonucléose chez les orthoptères. Annales de la Societé Entomologique de France 13:487-493.

5. 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.

6. Kapelinskaya TV, Martynova EU, Schal C, Mukha DV. 2011. Expression strategy of densonucleosis virus from the German cockroach, *Blattella germanica*. J Virol 85:11855-11870.

7. Huynh OT, Pham HT, Yu Q, Tijssen P. 2012. *Pseudoplusia includens* Densovirus Genome Organization and Expression Strategy. J Virol 86:13127-13128.

8. Pham HT, Huynh OTH, Jousset FX, Bergoin M, Tijssen P. 2013. Genome sequence and organization of *Junonia coenia* densovirus (JcDNV) Genome Announcements submitted.

9. Tijssen P, Li Y, El-Far M, Szelei J, Letarte M, Zadori Z. 2003. Organization and expression strategy of the ambisense genome of densonucleosis virus of *Galleria mellonella*. J Virol 77:10357-10365.

10. Huang X, Madan A. 1999. CAP3: A DNA sequence assembly program. Genome research 9:868-877.

11. 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.

12. Bergoin M, Tijssen P. 2010. Densoviruses : a highly diverse group of arthropod parvoviruses, p. 57-90. *In* Asgari S, Johnson KN (ed.), Insect Virology. Horizon Scientific Press, Norwich, U.K.

13. 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.

14. Fediere G, El-Far M, Li Y, Bergoin M, Tijssen P. 2004. Expression strategy of densonucleosis virus from *Mythimna loreyi*. Virology 320:181-189.

15. El-Mergawy R, Yi L, El-Sheikh M, El-Sayed M, Abol-Ela S, Bergoin M, Tijssen P, Fédière G. 2003. Epidemiology and biodiversity of the Densovirus MlDNV in the field populations of Spodoptera littoralis and other noctuid pests. Bulletin of Faculty of Agriculture, Cairo University 54:269-281.

16. Fédière G, Ul-Sheikh M, El Mergawy R, Salah M, Masri M, Abol-Ela S, Bergoin M, Tijssen P. 2003. Record of densoviral diseases occurring among the noctuid populations in lucerne fields at El-Bahareya oasis in Egypt. IOBC WPRS BULLETIN 26:237-240.

3.2.5. Relevance of our results on AdDNV to cricket industry and AdDNV molecular biology

AdDNV was first isolated in 1977 (Switzerland) and it is the first ambidensovirus discovered in the *Orthopteran* order of insects. The virus is highly virulent and specific to cricket, especially *A. domesticus*. Severe epidemics were happening in most cricket facilities in Europe in the last 35 years and in North America in 2009/2010. Most cricket samples from growers and pet stores were found to be infected by this virus. Subsequently, the native *Gryllus* species have been introduced in US to replace rearing *A.domesticus* crickets without strict regulatory supervision or even illegally. Our results showed that they are also suitable hosts for the virus albeit less permissive than the house cricket. Likewise, positive results obtained by PCR at highest dilutions (1000X) likely confirmed the fact that AdDNV replicated in at least *G. assimilis* and *G. locorojo* as well. *G. bimaculatus* was found to be positive to AdDNV but only at lowest (10-fold) or no dilution of the specimens. Cautions about these introduced species are therefore warranted since their spread can cause consequential impacts on local agriculture as well as native ecology. Assays for AdDNV presence were also positive in *G. signillatus*, a widely distributed cricket species that has been legally sold in US and new mutations have been identified by sequencing analysis (Publication in Annexe 10). Eight AdDNV isolates were obtained from different outbreaks in Europe (Germany; 2004, 2006, 2007 2009); North America (2009) and Japan (2012/2013) and from different cricket species, *Gryllodes* in Canada (GsCa12) and *G. locorojo* in the United States. NS1 and NS2 protein sequences between these isolates shared highest identity (99.3%) while identity was 94.4% for NS3. For two VP-ORFs, it was 98.1 and 97.1%, respectively. The Japanese and *G. sigillatus* isolates were together in the same clade and separated from others of Europe and North American strains (Figure 3-15), suggesting that these two strains had been influenced by other evolutionary pressures.

The expression strategy of AdDNV was studied in detail: mapping of viral transcripts and splicing sites by Northern blot, RACE, mass spectrometry and protein sequencing. Similar to Densovirus members, AdDNV potentially produced a set of 3 NS proteins NS1, NS2 and NS3. NS3 is produced from an unspliced transcript whereas NS1 and NS2 transcripts are spliced forms in which the NS3-ORF is completely removed. This leads to leaking scanning translation at the NS2 start codon since this region becomes more favorable than that of the upstream AUG of NS1. Expression of VPs involved unusual splicing features; two sets of generated VP proteins depends on the splicing of introns I (including Ia and Ib donors and I acceptor) versus

II. Splicing of these 2 introns was mutually exclusive in which, splicing of intron I (Ia and Ib) removed the donor site of intron II and resulted in VP2, VP3 and VP4 proteins. However, when splicing occurred for intron II that removed the acceptor I, two split ORFs were put in frame to generated VP1 proteins. This spliced form also gave rise to VP3 and VP4 probably by leaky scanning due to the strong initiation of VP4 start codon compared to that of VP3. The splicing characteristic for VP expression so far has not been reported in parvoviruses but has also been observed for BgDNV.

3.3. Section 3: Cloning of densoviral telomeres

3.3.1. Résumé en français de la stratégie et des travaux des séquences télomériques des génomes de parvovirus

Introduction

Comme nous l'avons mentionné précédemment, le clonage des séquences télomériques des génomes de parvovirus pose souvent des problèmes, notamment du fait de leur forte teneur en GC et de leurs séquences palindromiques permettant la formation de structures bicaténaires en épingle à cheveux. Or, ces extrémités jouent un rôle essentiel dans la réplication et aussi probablement dans leurs interactions avec les protéines de capside. En vue du clonage des séquences complètes de deux DNVs appartenant au genre *Densovirus*, le densovirus du lépidoptère *Pseudoplusia includens* (PiDNV) et le densovirus de *Junonia coenia* (JcDNV), nous avons mis au point une technique permettant de cloner leurs extrémités le plus efficacement possible. Cette technique repose sur la ligation d'une ancre d'ADN monocaténaire aux extrémités 3' des brins (+) et (–) du génome densoviral, à l'aide de l'ARN ligase du bactériophage T4. Après ligation, les extrémités sont amplifiées à partir d'un couple d'amorces, l'une complémentaire à l'ancre, l'autre d'une séquence virale spécifique. La PCR est réalisée en ajoutant de la bétaïne (1.3 M) du diméthylsufoxide (5%) et du 7-deaza-dGTP (50mM) dont le rôle est de minimiser la formation de structures secondaires et de favoriser la dénaturation de séquences d'ADN riches en GC.

Publication 5: Huynh OT, Pham HT, Yu Q, Tijssen P. J Virol. 2012 Dec; 86(23):13127-8. Le clonage du génome complet du PiDNV a été réalisé dans un premier temps dans le vecteur linéaire pJAZZ (Lucigen Corp), favorable au clonage de séquences instables, puis dans un vecteur circulaire. Afin de limiter les risques de délétions aux extrémités 5' et 3', les régions télomériques ont été également amplifiées en utilisant la technique des adaptateurs 3' terminaux décrite ci-dessus. Quatre clones complets ont été séquencés. Le génome du PiDNV a une taille de 5990 nucléotides incluant une répétition terminale inversée de 540 nt formant par repliement des 120 nt terminaux une structure bicaténaire typique en forme de Y. L'organisation des gènes de type ambisens est tout à fait semblable à celle des DNVs du genre *Densovirus* avec 3 ORFs sur la moitié 5' d'un brin codant pour les protéines NS1, NS2 et NS3 et un ORF occupant la même région sur le brin complémentaire codant pour les 4 protéines de capside. La position et la séquence des deux promoteurs codant les transcrits NS et VP ainsi que les

positions du site donneur et du site accepteur de l'épissage générant l'ARN messager codant pour NS1 et NS2 sont quasi identiques à celles déterminées chez les deux densovirus GmDNV et MlDNV qui ont fait l'objet d'une analyse précise de leur stratégie d'expression.

Publication 6: Pham HT, Huynh OT, Jousset FX, Bergoin M, Tijssen P. Genome Announc. 2013 Aug 8; 1(4). Le densovirus du lépidoptère *Junonia coenia* est le premier génome de parvovirus de type ambisens à être publié. Bien qu'infectieux, la séquence clonée du JcDNV dans le plasmide pBRJ était incomplète avec des délétions dans les répétitions terminales inversées. La séquence des régions codantes contenait d'autre part quelques ambiguïtés et erreurs conduisant notamment à une localisation erronée de l'AUG d'initiation de traduction de la protéine NS3. De plus, l'orientation des cadres de lecture codant pour les gènes NS et VP était inversée, avec les gènes VP à gauche (séquence 5' terminale) et les gènes NS à droite (séquence 3' terminale), contrairement à l'orientation de tous les génomes de parvovirus de vertébrés dont les gènes sont présentés par convention avec les NS à gauche et VP à droite. Afin de corriger ces anomalies, nous avons extrait l'ADN à partir d'un stock de virus qui avait servi à produire le pBRJ et que nous avait gracieusement fourni en 1983 le Dr. Tinsley de l'Université d'Oxford. Le grand fragment interne BamHI a été cloné dans pBluescript KS et les séquences ITRs amplifiées, puis clonées dans les mêmes conditions que celles décrites précédemment pour le virus PiDNV. Deux génomes complets ont été séquencés et leurs séquences comparées à celle de l'insert viral de pBRJ. Le génome complet du JcDNV a une taille de 6032 nucléotides (au lieu 5908 précédemment publié) dont deux répétitions terminales inversées de 545 nt de séquence quasi identiques. La séquence de l'ITR le plus long ne dépassait pas 517 nt dans pBRJ. Après correction des ambiguïtés et des erreurs de séquence dans les régions codantes, l'organisation des gènes codants pour les protéines NS et VP, la localisation des promoteurs et les sites d'épissage du transcrit NS sont identiques à ceux des autres membres du genre *Densovirus*.

3.3.2. Contributions of authors to Publications in Section 3

Publications 5 and 6: Oanh Huynh was the main investigator as part of her M.Sc. study on PiDNV. Hanh Pham and Peter FX Jousset made a construct of the original incomplete clone of JcDNV. Since there were many difficulties in the obtaining of hairpin sequences of these viruses, Hanh Pham developped an alternative method to amplify and clone the terminal sequences of PiDNV and JcDNV. Qian Yu participated in the cloning steps. Pham and Tijssen prepared the JcDNV manuscript. M. Bergoin and P. Tijssen supervised the projects.

3.3.3. Publication 5: *Pseudoplusia includens* densovirus genome organization and expression strategy

J Virol. 2012 Dec; 86(23):13127-8

Huynh OT, Pham HT, Yu Q, Tijssen P.

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

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.

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) Densoviruses 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 pJazz 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 (MlDNV) (4). 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 (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.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 looplure-baited traps. J. Econ. Entomol. 75:647– 650.

2. Chao Y-C, Young SY III, Kim KS, Scott HA. 1985. A newly isolated densonucleosis virus from *Pseudoplusia includens* (Lepidoptera: Noctu-idae). 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* densovirus 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 densonucleosis 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.

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. Hod-der Arnold, London, United Kingdom

11. Tijssen P, et al. 2003. Organization and expression strategy of the am-bisense 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 denso-virus 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.

3.3.4. Publication 6: *Junonia coenia* densovirus (JcDNV) genome structure

Genome Announc. 2013 Aug 8; 1(4)

Pham HT[1], Huynh OT[1], Jousset FX[2], Bergoin M[2], Tijssen P[1].

[1]INRS, Institut Armand-Frappier, Laval, QC, Canada

[2]Laboratoire de Pathologie Comparée , Université Montpellier 2, Montpellier 34095, France

Abstract

The sequence of *Junoniae coenia* densovirus was the first genome sequence published among densoviruses, but the first published sequence contained incomplete inverted terminal repeats and ambiguous nucleotides or indels leading to an incorrect map of the open reading frames. Our sequencing of clones of the complete genome demonstrated that this virus is closely related to other viruses in the *Densovirus* genus.

The common buckeye (*Junonia coenia* [Hübn.]) is a butterfly in the *Nymphalidae* family. A nonoccluded, small virus was isolated from caterpillar cadavers (1) and subsequently characterized as a densovirus (2). This *Junonia coenia* densovirus (JcDNV) was kindly provided to us by T. W. Tinsley (NERC Institute of Virology, Oxford). Its restriction map showed a close relatedness to the GmDNV densovirus from *Galleria mellonella* (3). An infectious JcDNV genome was cloned (pBRJ) and sequenced (4, 5). Related densoviruses have been isolated, cloned, and sequenced, e.g., from *Galleria mellonella* [GmDNV (6)], *Mythimna loreyi* [MlDNV (7)], *Helicoverpa armigera* [HaDNV (8)], and *Pseudoplusia includens* [PiDNV (9)]. Compared to the published genome sequences of this group, the published JcDNV sequence had a different genome orientation, incomplete inverted terminal repeats (ITRs), and a different map of open reading frames (ORFs). In addition, there are several ambiguous sequences in the reported genome sequence of JcDNV (GenBank accession num-ber NC_004284).

Here, the entire JcDNV genome was extracted from the same virus stock as that used to produce pBRJ, recloned, and sequenced and compared to the genome sequence of pBRJ. The separately encapsidated, complementary DNA strands reannealed upon extraction. The central 5.45-kb part of the genome sequence, after BamHI digestion, was cloned into pBluescript KS+. The rest of the inverted terminal repeats (ITRs) were obtained by PCR from gel-purified 2.4- and 3.6-kb dsDNA fragments after digestion with NdeI. DNA was heated up to 95°C for 5 min, and 10 ng of each fragment was used for ligation with a 5P-ACGCAAGTACCGTGGTACCATGGATCCGG-3C3 adapter, including 1% DMSO, 1 mM hexamine cobalt chloride, and 10% polyethylene glycol (PEG) (final concentrations) and T4 RNA ligase (NEB), and then incubated overnight at room temperature. The DNA was then precipitated and eluted in 70 μl of sterile water. Five microliters was used for 25 μl of PCR, including 6% DMSO, 1.3 M betaine, 50 mM 7-deaza-dGTP, CTTCGGATCCTCT CCATCATC, and CCGGATCCATGGTACCACGGTACTT-GCGT as specific and adapter primers, respectively, and Phusion High-Fidelity DNA polymerase (98°C, 3 min; 25 cycles of 98°C 10 s, 65°C 20 s, and 72°C 20 s; elongation, 72°C, 5 min). Amplicons were cloned into a pGEMT-easy TA vector (Promega) and transformed into Sure cells. Two complete clones were sequenced in both directions and pBRJ was sequenced at locations of discrepancies, using Sanger's method and the primer-walking method as described before (6). The contigs were assembled by using the CAP3 program (http://pbil.univ-lyon1.fr/cap3.php/) (10).

The 6,032-nucleotide (nt) JcDNV genome contained 547-nt-long, nearly identical ITRs, the only differences being 396G, 537G, and the two TATA boxes (539-TATAAAT for the NS-

gene cassette and 5488-TATATAA for the VP-gene cassette). The typical terminal Y-shaped hairpins of 130 nt contained two orientations at sequences 51 to 80 and 5953 to 5982, "flip" and its reverse complement orientation "flop." The ambisense JcDNV genome sequence was 83 to 87% identical to those of other viruses in the *Densovirus* genus. The ORFs, in contrast to the previous entry in GenBank, were conserved with members of the *Densovirus* genus, as were the splicing sites (11) with those identified for GmDNV (6) and MIDNV (7). Also, sequencing errors, gaps, and ambiguous nts were corrected and the ITRs completed. JcDNV contained the typical NS-1 helicase superfamily III and VP phospholipase A2 (12) motifs observed in other parvoviruses.

Nucleotide sequence accession number. The GenBank accession number of JcDNV is KC883978.

Acknowledgments

This work was supported by funding from the Natural Sciences and Engineering Research Council of Canada to P.T.; H.T.P. and O.T.H.H. acknowledge tuition fee waivers at INRS.

References

1. Rivers CF, Longworth JF. 1972. A nonoccluded virus of *Junonia coenia* (*Nymphalidae*: *Lepidoptera*). J. Invertebr. Pathol. 20:369–370.
2. Longworth JF. 1978. Small isometric viruses of invertebrates. Adv. Virus Res. 23:103–157.
3. Jousset FX, Jourdan M, Compagnon B, Mialhe E, Veyrunes JC, Bergoin M. 1990. Restriction maps and sequence homologies of two densovirus genomes. J. Gen. Virol. 71(Pt 10):2463–2466.
4. Jourdan M, Jousset FX, 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 179:403–409.
5. Dumas B, Jourdan M, Pascaud AM, Bergoin M. 1992. Complete nucle-otide sequence of the cloned infectious genome of *Junonia coenia* densovirus reveals an organization unique among parvoviruses. Virology 191: 202–222.

6. 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.

7. Fédière G, El-Far M, Li Y, Bergoin M, Tijssen P. 2004. Expression strategy of densonucleosis virus from *Mythimna loreyi*. Virology 320: 181–189.

8. 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. doi: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. Huang X, Madan A. 1999. CAP3: a DNA sequence assembly program. Genome Res. 9:868–877.

11. Wang Y, Abd-Alla AM, Bossin H, Li Y, Bergoin M. 2013. Analysis of the transcription strategy of the *Junonia coenia* densovirus (JcDNV) genome. Virus Res. 174:101–107.

12. Zádori Z, Szelei J, Lacoste MC, Li Y, Gariépy S, Raymond P, Allaire M, Nabi IR, Tijssen P. 2001. A viral phospholipase A_2 is required for parvovirus infectivity. Dev. Cell 1:291–302.

3.3.5. Schematic overview of telomeric cloning

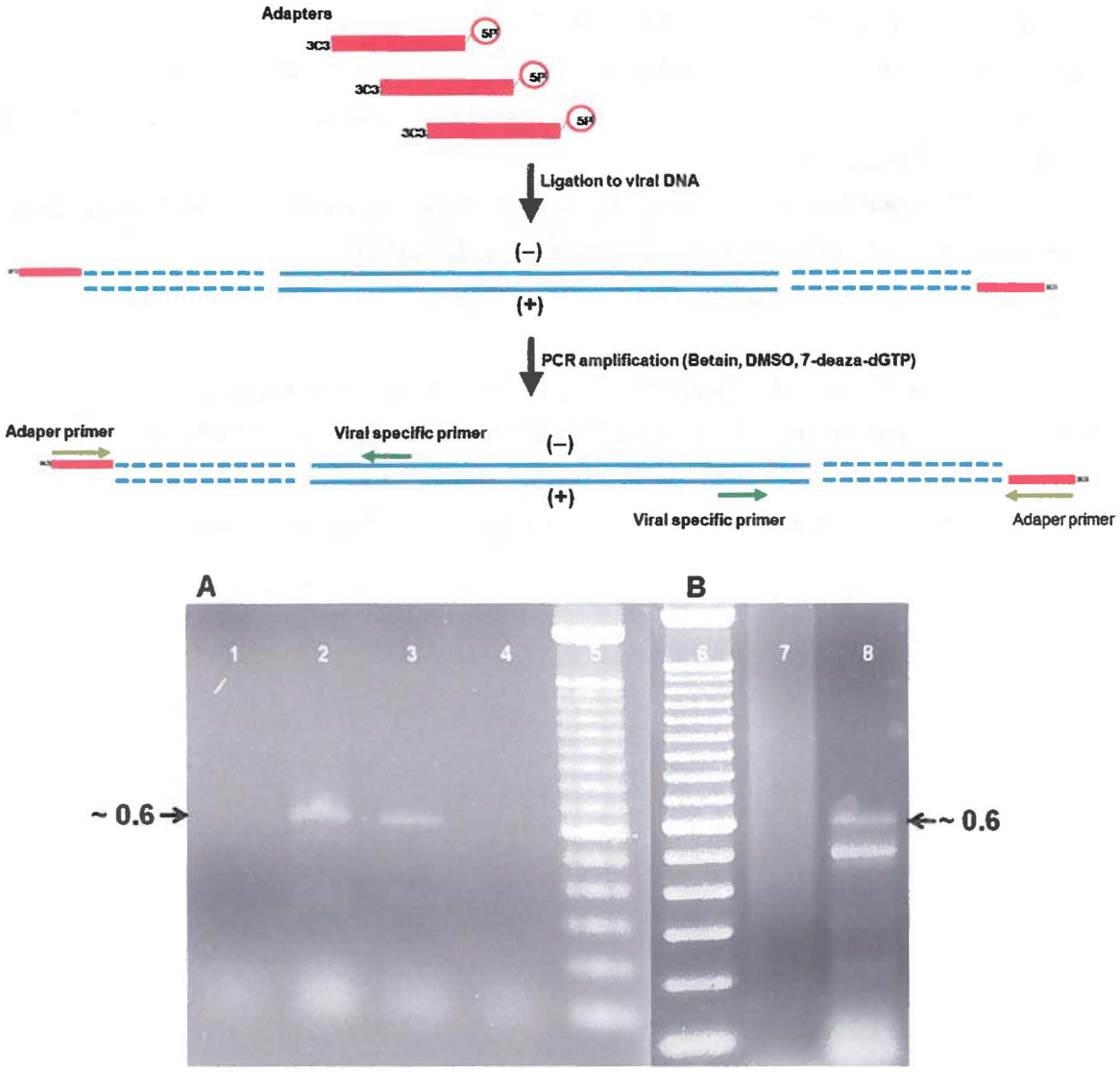


Figure 3-16. Adapter ligation-mediated PCR method to amplify unknown sequences of parvoviral telomeres. Adapters are single-stranded oligos with 5' phosphorylated- and 3' blocked- ends that can be ligated into viral DNA by T4 RNA ligase. Results of PCR amplification of PiDNV(A) and JcDNV ITRs (B) using adapter primer and viral specific primer in the addition of Betain, DMSO and 7-deaza-dGTP. Lane 1, 2, 3 and 4: different viral DNA concentrations input in ligation reactions, 100, 10, 1 and 0.1 ng, respectively. Lane 5 and 6: 1kb ladder (Invitrogen). Lane 7 and 8: 100 ng and 10 ng of DNA template.

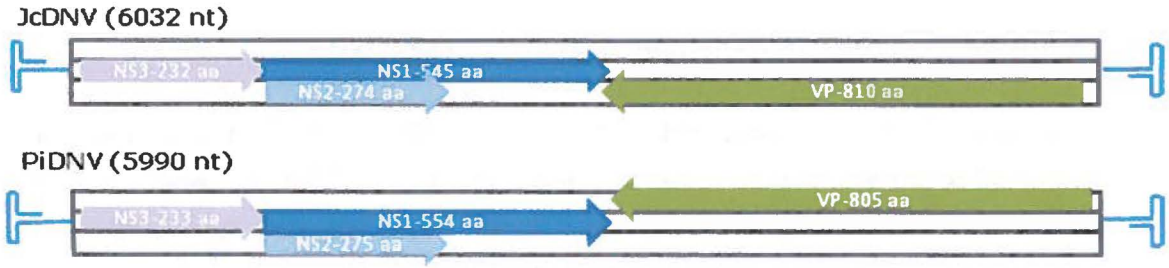


Figure 3-17. Similar enomic structure of JcDNV and PiDNV: ambisense genome encode for NS1, NS2 and NS3 proteins on left half and VP on right half; ITRs contain Y-shaped hairpins.

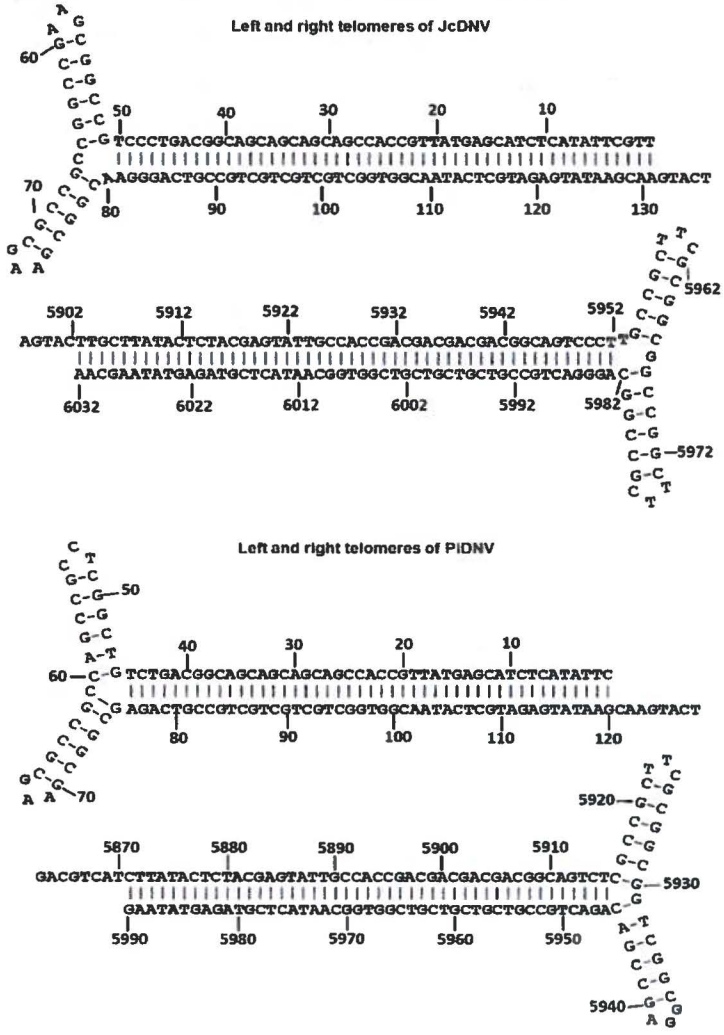


Figure 3-18. Nucleotide sequence and Y-shaped structure of telomeres in JcDNV (130 nts, 61% GC) and PiDNV (120 nts, 61% GC).

3.3.6. Impact of telomeric cloning on parvovirus research

Parvoviral telomeres have critical functions in DNA replication and may contain transcriptional control elements or sequences responsible for viral DNA and capsid interactions. Therefore, obtaining entire genomes of parvoviruses is an important step for subsequent biological studies. Here, we have developed an alternative method for amplification and cloning of parvoviral termini. This method allows determining unknown hairpin sequence of densovirus specimens with low virus content at least at 1 ng/µl. This is a combination of ligation and PCR amplification steps. First, the 3'-end of plus and minus DNA strands were ligated with a known adapter, and then PCR was followed using adapter and viral specific primers. Limitation by low viral DNA contents, secondary structure, and high GC content could be eliminated by PCR accompanied with betain, DMSO, and 7-deaza-dGTP. By this method, large ITRs of PiDNV and JcDNV have been obtained containing 540 nts and 547 nts, respectively. Both viruses have stable Y-shaped hairpin structures at their ends. Amplified products from ligation products containing 10 and 1 ng viral DNA input were visible on agarose gels whereas increasing amounts of DNA input caused smears or no clear bands after PCR. The 5' phosphorylated- and 3' blocked- adapters at high concentration were used in order to permit the ligation between adapter and viral DNA and prevent self-ligation of DNA template. Ligation reactions were optimized in 10% PEG and 1% DMSO final concentrations to destabilize the internal secondary structure of viral telomeres and increase interactions between T4 RNA ligase and DNA molecules. The purity level of DNA, the stability of hairpin and the GC content of viral telomeres were important factors for the efficacy of this method.

3.4. Section 4: Novel single-stranded DNA viruses that infect crickets and shrimps

3.4.1. Résumé en français de la découverte de nouveaux virus dans les élevages industriels de crevettes et de grillons.

Introduction

A l'occasion de l'enquête épidémiologique sur les épizooties à AdDNV survenues dans les élevages de masse de grillons de différentes régions du globe (Publication in Annexe 10), la recherche du virus par PCR s'est avérée négative dans certains échantillons. De même, la recherche de PstDNV dans des échantillons de crevettes provenant du Vietnam a parfois été négative. Ces résultats nous ont conduits à rechercher si les mortalités constatées dans ces élevages n'étaient pas dues à d'autres types de virus. Comme l'illustrent les 5 publications de ce chapitre, ces investigations ont conduit à la découverte de plusieurs nouveaux petits virus à ADN chez ces deux familles d'invertébrés. Ainsi les Publications 7, 8 et 9 rapportent la découverte chez les grillons et chez la crevette *Penaeus monodon* de virus dont le génome est une molécule d'ADN circulaire monocaténaire de petite taille les rapprochant aux virus de la famille *Circoviridae,* mais avec lesquels ils ne présentent que très peu d'homologies de séquence. Deux autres virus ont été découverts chez *A. domesticus*, un densovirus dont le génome de 5 kb seulement possède une organisation des gènes de type ambisens semblable à celle connue chez les membres du genre *Densovirus* avec lesquels il ne présente toutefois qu'une très faible similarité de séquence (Publication 10) et un densovirus de taille et de structure génomique de type brevidensovirus mais ne possédant pas le gène VP (Publication 11).

Publication 7: Pham HT, Bergoin M, Tijssen P. Genome Announc. 2013 Mar 14; 1(Hsu *et al.*). **Le volvovirus d'*Acheta domesticus*, un nouveau virus à ADN simple brin circulaire.**

Parmi les nombreux échantillons de cadavres de grillons *A.domesticus* provenant de differents élevages d'Amérique du Nord, certains étaient négatifs à l'AdDNV par PCR. La recherche d'un autre virus dans les broyats d'un de ces échantillons a révélé la présence, dans les culots d'ultracentrifugation, de particules virales icosaédriques de taille légèrement plus petite que les densovirus (18 nm vs 22 nm) (Figure 3-19C). La très forte concentration en virus dans ces culots implique une multiplication importante du virus chez le grillon et peut expliquer

les fortes mortalités observées dans les élevages. L'ADN extrait de ce virus étant réfractaire à la coupure par enzymes de restriction, nous avons donc déduit qu'il était vraisemblablement monocaténaire. Après synthèse du brin complémentaire par l'ADN polymérase du bactériophage ø29, le génome viral digéré par MboI a été cloné et séquencé. L'analyse de la séquence a révélé qu'il s'agit d'une molécule d'ADN circulaire de 2517 nucléotides possédant un site unique EcoRI. La nature circulaire du génome viral et sa taille ont été confirmées par PCR inverse à partir d'amorces prises de part et d'autre du site unique de restriction EcoRI. La détection sur la séquence du nonanucléotide TAGTATTAC formant une structure bicaténaire de type tige-boucle caractéristique de l'origine de réplication des génomes de circovirus et des cyclovirus a permis de proposer cette séquence comme l'origine de réplication du virus et d'orienter le génome (Figure 3-19B). Quatre cadres de lecture ouverts (ORFs) localisés de part et d'autre de cette séquence sont présents sur la molécule: deux dans la direction sens (ORF1 et ORF4) et deux dans la direction antisens (ORF2 et ORF3) (Figure 3-19A). La recherche d'homologies nucléotidiques (Blastn) avec les séquences de GenBank n'a donné aucun résultat. Par contre, la recherche d'homologies au niveau des séquences protéiques (Blastp) a révélé uniquement 30% d'identité entre l'ORF2 et les protéines de réplication des circovirus et des cyclovirus. La structure circulaire de ce génome, ses très faibles homologies avec les virus de la famille *Circoviridae* et les différences de taille et d'organisation des gènes nous ont conduits à proposer le nom de Volvovirus d'*Acheta domesticus* (AdVVV) comme nouveau genre dans la famille *Circoviridae*.

Publication 8: Pham HT, Iwao H, Bergoin M, Tijssen P. Genome Announc. 2013 Jun 27; 1(3). **Nouveaux isolats de volvovirus d'*Acheta domesticus* (Japon) et *Gryllus assimilis* (États-Unis).**

Suite à la découverte du premier volvovirus d'*Acheta domesticus*, nous avons retrouvé dans les broyats de deux autres échantillons, l'un *d'A. domesticus* provenant du Japon, l'autre de grillons du genre *Gryllus assimilis* provenant d'un élevage en masse aux États-Unis, des virions de 18 nm de diamètre. Après purification, le génome de ces virus a été extrait, cloné et séquencé dans les mêmes conditions que celles décrites dans la Publication 7. Le génome de ces virus est un ADN monocaténaire circulaire de 2517 nt pour l'isolat japonais (AdVVV-Japan) et de 2516 nt pour l'isolat de *Gryllus assimilis* (AdVVV-Ga). Les deux virus présentent la même organisation de leurs gènes que le premier AdVVV décrit (AdVVV-IAF). Les différences dans la séquence concernent 22 substitutions (aucune dans les ORF2 et ORF3 se chevauchant) dans la séquence de l'ORF1 et une dans la région intergénique. L'essentiel des substitutions (18 dont

14 non-synonymes) sont dans l'ORF1 présumée codée pour la capside tandis que 3 substitutions non synonymiques se trouvent dans la protéine présumée codée par l'ORF4. Les substitutions non synonymiques (4) dans le génome de l'isolat AdVVV-Ga sont toutes dans la protéine de capside et aux mêmes positions que celles de l'isolat AdVVV-Japan. Ces résultats montrent la vaste dispersion géographique des volvovirus et leur capacité d'infecter plusieurs hôtes.

Publication 9: Pham HT, Yu Q, Boisvert M, Van HT, Bergoin M, Tijssen P. Genome Announc. 2014 Jan 16; 2(1). **Un virus de type circovirus isolé de la crevette *Penaeus monodon*.**

L'examen au microscope électronique de broyats de crevettes du genre *P. monodon* d'origine vietnamienne était faiblement positif à la détection du PstDNV et a révélé la présence de particules virales de taille comparable à celle du volvovirus d'*A. domesticus* dans l'échantillon VN11. Ce virus a été purifié, son génome extrait, cloné et séquencé comme indiqué dans la Publication 7. Il s'agit d'une molécule d'ADN monocaténaire circulaire de 1777 nucléotides, présentant la structure tige-boucle typique de l'origine de réplication et une organisation des gènes semblables à celles des virus de la famille *Circoviridae*. Trois cadres de lecture (ORFs) majeurs sont présents sur cette molécule. Le plus grand codant pour une protéine de 266 résidus d'acides aminés présente 30% d'identité avec la protéine Rep des circovirus. Dans la direction opposée, un ORF de 255 aa présente 25% d'identité avec le gène de capside du circovirus d'une espèce de copépode du genre *Diporeia*. Le troisième ORF de 146 aa ne présente aucune homologie de séquence significative avec les protéines de GenBank. Ce virus a été nommé PmCV-1 (Figure 3-20A).

Un second virus de ce type a été détecté dans un autre échantillon de crevettes d'origine vietnamienne. L'analyse de son génome de 1788 nucléotides ne révèle aucun ORF de taille significative en utilisant les codes génétiques standards. Par contre, le code «Ciliate, Dasycladacean and Hexamita Nuclear» a permis de détecter trois ORFs majeurs, tous orientés dans la même direction. Le plus grand ORF de 276 aa code pour une protéine présentant 42% d'homologie avec la protéine Rep du circovirus du copépode *Labidocera aestiva*. Le deuxième, situé en aval des deux premiers, a une taille de 223 aa et code probablement pour une protéine de capside. Le troisième ORF, 134 aa chevauche complètement sur la séquence du premier ORF mais ne présente aucune identité avec une séquence dans la banque de protéines de

GenBank (Figure 3-20B). Les promoteurs de ces gènes montrent certaines propriétés communes avec ceux des gènes de procaryotes. Ce virus a été nommé PmaCV-2 (*Peneaus monodon*-associated circovirus-2). L'ensemble de ces résultats apporte la preuve de la grande diversité des petits virus à ADN de type CRESS. Nos résultats permettent d'étendre le spectre d'hôtes de ce groupe émergent de virus à de nouvelles espèces d'invertébrés terrestres et marins.

Publication 10: Pham H.T., Yu Q., Bergoin M., Tijssen P. 2013 Genome Announc. 2013 1(6): e00914-13. **Un nouveau type de densovirus ambisens, le Mini Ambidensovirus du grillon *Acheta domesticus*.**

Nous avons reçu, au début de l'année 2013, d'une ferme américaine à *A. domesticus* où une épizootie s'était déclarée, un échantillon de cadavres de grillons dans lequel la recherche de virus de type AdDNV ou AdVVV s'est avérée négative. L'observation dans les broyats de ces grillons de très nombreux virus de 23 nm de diamètre (Figure 3-21B) nous a conduit à purifier ce virus et à en extraire un ADN dont l'analyse en gel d'agarose a permis d'estimer sa taille à environ 5kb (Figure 3-21A). Nous avons alors cloné l'ADN viral après remplissage des extrémités à l'aide du grand fragment de Klenow et de l'ADN polymérase de T4 au site EcoRV du plasmide pBluescriptSK(+). Le clonage de la région 5' terminale du génome a été facilité par la présence dans la séquence virale d'un site unique EcoRV dans cette région. La séquence complète du génome est de 4945 nt dont la structure et l'organisation des gènes ressemblent, en plus court, à celle des génomes du genre *Densovirus* mais dont la similarité de séquence avec celles des densovirus est très faible. On trouve aux extrémités 3' et 5', des séquences non codantes de 282 nt et 233 nt dont les 199 nt terminaux sont des ITRs. Par repliement, les 113 nt des extrémités des ITRs forment des structures bicaténaires en forme d'Y (Figure 3-22). L'organisation des gènes est de type ambisens avec les trois ORFs codant pour les 3 protéines non-structurales localisées sur la moitié 5' d'un brin, dans l'ordre NS3, NS1 et NS2 (chevauchant sur NS1). La similarité de séquence la plus forte concerne la protéine NS1 (25% de similarité avec les NS1 d'iteravirus) dans laquelle on retrouve les motifs RCR et hélicase typiques des NS1 de parvovirus. Le plus grand ORF occupe la moitié 5'-terminale du brin complémentaire et code pour les protéines de capside. Il porte dans la région N-terminale de VP1 le motif PLA2 très conservé des parvovirus. Trois ATG en aval et en phase avec l'ATG de VP1 pourraient servir à la synthèse de 3 autres protéines de capside par leaky scanning comme c'est la règle chez les densovirus. Les plus fortes similarités de séquence de l'ORF VP sont de

l'ordre de 25% avec les VP des membres du genre *Densovirus*. Deux séquences TATA box sont présentes jouxtant les séquences ITRs, l'une en amont de l'ORF NS3, l'autre en amont de l'ORF VP (Figure 3-23). Comme dans le genre *Densovirus*, un site accepteur potentiel a été détecté juste en amont de l'ATG de NS1 permettant l'épissage alternatif du transcrit NS par élimination de la séquence NS3. Il ressort de l'ensemble des données fournies par l'analyse du génome de ce virus que l'on a affaire à un nouveau genre de densovirus ambisens pour lequel nous proposons le nom d'AdMADV pour *A.domesticus* Mini Ambidensovirus.

Publication 11 : Pham HT *et al.*, en préparation. **Un virus de type brevidensovirus isolé d'*A. domesticus* suite à des mortalités importantes dans un élevage au Canada**

Nous avons isolé récemment un virus de petite taille dans des lots de grillon domestique provenant d'une ferme de l'Ontario où des mortalités anormales se produisent épisodiquement. Le virus, très abondant dans les cadavres, a été purifié et son génome extrait selon les protocoles décrits dans les publications précédentes. Après traitement de l'ADN viral à l'ADN polymérase du bactériophage ø29 ou l'ADN polymérase du phage T4, sa taille mesurée en gel d'agarose est légèrement supérieure à 3 kb. Les fragments générés par digestion par MboI ont été clonés et séquencés et la taille du génome déterminée à 3233 nucléotides. L'analyse de cette séquence révèle qu'elle renferme deux ORFs majeurs. L'un de 795 aa présente des similarités de 38% avec la protéine NS1 du densovirus de *Culex pipiens* et possède les motifs RCR et hélicase des NS1 de tous les parvovirus. La séquence du deuxième ORF chevauche complètement celle de NS1 mais dans un autre cadre de lecture. Cet ORF présente 40% d'homologie avec l'ORF NS2 du brevidensovirus AalDNV. Les extrémités 5' et 3' du génome forment par repliement des structures en épingle à cheveux asymétriques comparables à celles des brevidensovirus. La propriété la plus remarquable de cette séquence est qu'elle ne renferme aucun cadre de lecture codant pour une protéine de capside. Deux hypothèses sont proposées pour expliquer la formation de ces virus: soit il existe un second segment de ce virus portant le gène VP, comme c'est le cas pour les bidensovirus, soit le virus utilise la capside d'un autre virus comme le suggère la présence fréquente de l'AdVVV dans les broyats.

3.4.2. Contribution of authors to publications in Section 4

Dr. H. Iwao (Japan) sent us cricket samples from Niigata that contained volvovirus. Hanh Thi Van (Vietnam) sent us shrimps that contained a circo-like virus. Hanh Pham was

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

3.4.3. New circular single-stranded virus from different cricket species and different regions

3.4.3.1. Publication 7: *Acheta domesticus* volvovirus, a novel single-stranded circular DNA virus of the house cricket

Genome Announc. 2013 Mar 14; 1(2)

Pham HT, Bergoin M, Tijssen P

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

Abstract

The genome of a novel virus of the house cricket consists of a 2,517-nucleotide (nt) circular single-stranded DNA (ssDNA) molecule with 4 open reading frames (ORFs). One ORF had a low identity to circovirus nucleotide sequences (NS). The unique properties of this volvovirus suggested that it belongs to a new virus family or genus.

Cricket-breeding facilities in the United States produce billions of pet-feeder crickets annually (1, 2). The preferred house cricket, *Acheta domesticus*, is highly susceptible to a densovirus, *Acheta domesticus* densovirus (AdDNV), which has caused severe outbreaks since September 2009 and decimated *A. domesticus* stocks in North America. Samples received from die-offs were invariably positive for this virus. However, some recently received samples from mass cricket die-offs in North America were negative for AdDNV.

AdDNV-negative crickets (20 g) were homogenized in 20 ml of a 3:1 mixture of phosphate-buffered saline (PBS) and carbon tetrachloride. After low-speed centrifugation, the upper aqueous phase was passed through 0.45 nm filters and putative viruses were pelleted by centrifugation for 1.5 h at 40,000 rpm and resuspended in a small volume of Tris-EDTA (TE) buffer followed by DNase A and RNase A treatments to remove contaminating nucleic acids. Electron microscopy examination of a 100-fold dilution of the resuspended pellet revealed highly concentrated icosa-hedral particles of about 18 nm in diameter.

DNA extracted from purified virus by the High Pure viral nucleic acid kit (Roche Applied Science) was resistant to restriction endonucleases and presumably single stranded. Native viral DNA was used for double-stranded DNA synthesis at 30°C by 29 DNA polymerase (3). Amplified DNA was digested with MboI, cloned into the BamHI site of the pBluescriptSK(–) vector, and sequenced by Sanger's method and primer walking as described before (4). The sequences were assembled by the CAP3 program (5) and generated a 2,517-nucleotide (nt) sequence containing a single EcoRI site. PCR using native DNA and 2 sets of outward primers (with respect to the EcoRI fragment) and sequencing confirmed the circular nature of the genome and the size of 2,517 nt. Due to the circular (rolling) nature of the genome, the name *Acheta domesticus* volvovirus (AdVVV; Volvo [Latin] roll) was proposed.

Numbering of the genome started with the putative nonanucleotide origin of replication (1-TAGTATTAC), located, as for circovirus or cycloviruses (6), between the open reading frames (ORFs) with opposite orientations. Among ORF products of 100 amino acids (aa), ORF1 (361 aa, starting at nt 447) and ORF4 (130 aa, starting at nt 70) were in the sense direction, whereas ORF2 (270 aa, starting at nt 2445) and the overlapping ORF3 (207 aa, starting at nt 2393) were in the antisense direction. BLASTn failed to detect any identity to viral sequences. However, BLASTp revealed a maximum identity of about 30% between ORF2 and Rep proteins of circoviruses and cycloviruses, with a coverage of ~85% (aa 5 to 80, Viral_Rep superfamily [pfam02407], and aa 150 to 212, P-loop_NTPase [pfam00910]). The other ORFs did not have any viral identity using BLASTp.

The lack of sequence identity and the differences in genome organization and size indicated a new virus family or genus. To our knowledge, this is the first circular single-stranded DNA virus in insects that is not related to cycloviruses (7, 8), circoviruses (9–11), nanoviruses (12, 13), or geminiviruses (14, 15), and it may be of interest in elucidating the evolution of this rapidly expanding virus group.

Nucleotide sequence accession number. The GenBank accession number for AdVVV is KC543331.

Acknowledgments

This work was supported by funds from the Natural Sciences and Engineering Research Council of Canada to P.T.; H.T.P. acknowledges tuition waivers at INRS.

References

1. 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.
2. 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 p-tential disaster. Zootaxa 3504:67–88.
3. de Vega M, Lázaro JM, Mencía M, Blanco L, Salas M. 2010. Im-provement of 29 DNA polymerase amplification performance by fusion of DNA binding motifs. Proc. Natl. Acad. Sci. U. S. A. 107: 16506–16511.
4. 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.
5. Huang X, Madan A. 1999. CAP3: a DNA sequence assembly program. Genome Res. 9:868–877.
6. 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.

7. Rosario K, Dayaram A, Marinov M, Ware J, Kraberger S, Stainton D, Breitbart M, Varsani A. 2012. Diverse circular ssDNA viruses discovered in dragonflies (Odonata: Epiprocta). J. Gen. Virol. 93:2668–2681.

8. Rosario K, Marinov M, Stainton D, Kraberger S, Wiltshire EJ, Collings DA, Walters M, Martin DP, Breitbart M, Varsani A. 2011. Dragonfly cyclovirus, a novel single-stranded DNA virus discovered in dragonflies (Odonata: Anisoptera). J. Gen. Virol. 92:1302–1308.

9. 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.

10. Johne R, Fernández-de-Luco D, Höfle U, Müller H. 2006. Genome of a novel circovirus of starlings, amplified by multiply primed rolling-circle amplification. J. Gen. Virol. 87:1189–1195.

11. Kapoor A, Dubovi EJ, Henriquez-Rivera JA, Lipkin WI. 2012. Complete genome sequence of the first canine circovirus. J. Virol. 86:7018.

12. Timchenko T, Katul L, Aronson M, Vega-Arreguín JC, Ramirez BC, Vetten HJ, Gronenborn B. 2006. Infectivity of nanovirus DNAs: induction of disease by cloned genome components of Faba bean necrotic yellows virus. J. Gen. Virol. 87:1735–1743.

13. Timchenko T, Katul L, Sano Y, de Kouchkovsky F, Vetten HJ, Gronenborn B. 2000. The master rep concept in nanovirus replication: identification of missing genome components and potential for natural genetic reassortment. Virology 274:189–195.

14. Buchmann RC, Asad S, Wolf JN, Mohannath G, Bisaro DM. 2009. Geminivirus AL2 and L2 proteins suppress transcriptional gene silencing and cause genome-wide reductions in cytosine methylation. J. Virol. 83: 5005–5013.

15. Martin DP, Lefeuvre P, Varsani A, Hoareau M, Semegni JY, Dijoux B, Vincent C, Reynaud B, Lett JM. 2011. Complex recombination patterns arising during geminivirus coinfections preserve and demarcate biologi-cally important intra-genome interaction networks. PLoS Pathog. 7:e1002203. doi:10.1371/journal.ppat.1002203

3.4.3.2. Publication 8: New volvovirus isolates *from Acheta domesticus* (Japan) and *Gryllus assimilis* (United States)

Genome Announc. 2013 Jun 27; 1(3)

Pham HT[1], Iwao H[2], Bergoin M[1], Tijssen P[1]

[1] INRS, Institut Armand-Frappier, Laval, QC, Canada

[2] Marinpia Nihonkai 5932-445, Nishifunami, Niigata, Japan

Abstract

A novel circular single-stranded DNA (ssDNA) virus, volvovirus, from the house cricket has been described recently. Here, we report the isolation of volvoviruses from *Acheta domesticus* in Japan and *Gryllus assimilis* in the United States. These *Acheta domesticus* volvovirus (AdVVV) isolates have genomes of 2,517 and 2,516 nucleotides (nt) and 4 large open reading frames (ORFs).

Billions of pet-feeder crickets are produced annually (1, 2). Since September 2009, stocks of the preferred house cricket, *Acheta domesticus*, were decimated due to severe fatal outbreaks caused by a densovirus, the *Acheta domesticus* densovirus (AdDNV) (2, 3). Similarly, in Japan house cricket stocks were decimated since early in the summer of 2009 despite strict import and export regulations. To avoid heavy losses, many cricket producers switched to the Jamaican field cricket, *Gryllus assimilis* (Fabricius) (1). Some recent samples contained a new single-stranded, circular DNA virus (4), volvovirus or *Acheta domesticus* volvovirus (AdVVV)-IAF, that is not related to cycloviruses (5, 6), circoviruses (7, 8, 9), nanoviruses (10, 11), or geminiviruses (12, 13, 14).

Electron microscopy of diluted extracts from recently obtained dead house crickets from Japan and dead Jamaican field crickets from U.S. producers revealed highly concentrated icosahedral particles of about 18 nm diameter. Virus and nucleic acid were extracted as described previously (4). The nucleic acid was resistant to RNase and restriction enzymes, suggesting a single-stranded DNA (ssDNA) genome. Native viral DNA was used for double strand DNA synthesis at 30°C by phi29 DNA polymerase (3, 15). Amplified DNA was digested with MboI or EcoRI, cloned into the BamHI or EcoRI sites of the pBluescriptSK(-) vector, and sequenced by Sanger's method and primer walking as described before (4, 16). The sequences were assembled by the CAP3 program (17) and generated a 2,517-nucleotide (nt) sequence for the Japanese isolate (AdVVV-Japan) and a 2,516-nt sequence for the *Gryllus assimilis* isolate (AdVVV-Ga). Both were closely related to AdVVV-IAF (4) and contained a single EcoRI site. PCR using native viral DNA and 2 sets of outward primers (with respect to the EcoRI fragment) and sequencing confirmed the circular nature of the genome. Nucleotide numbering was as for AdVVV-IAF (4).

Among open reading frames (ORFs) coding for 100 amino acids (aa), ORF1 (361 aa, starting at nt 447), putatively coding for the capsid protein (CP), and ORF4 (130 aa, starting at nt 70) were in the sense direction, whereas ORF2 (270 aa, starting at nt 2445 for AdVVV-Japan but at nt 2444 for AdVVV-Ga) and the overlapping ORF3 (207 aa, starting at nt 2393 for AdVVV-Japan but at 2392 for AdVVV-Ga) were in the antisense direction. BLASTp revealed a maximum identity of about 30% between ORF2 and rep proteins of circoviruses and cycloviruses, with coverage of ~85%. Compared to the original AdVVV-IAF isolate, AdVVV-Japan contained 22 substitutions of which none was in the over-lapping ORF2 and ORF3, and only one was in the intergenic regions. Surprisingly, the bulk of substitutions were in the putative CP (18, of which 14 were nonsynonymous or 4% of protein sequence), whereas the putative

ORF4 protein contained 3 nonsynonymous substitutions. Also striking was the high arginine content of the N terminus of the CP. The AdVVV-Ga isolate, compared to AdVVV-IAF, contained 4 nonsynonymous substitutions in the capsid protein also found in the Japanese isolate and one additional nonsynonymous substitution. It also had a deleted T at position 1620 in the intergenic region.

In conclusion, volvoviruses seem to be widespread and to infect several species of crickets.

Nucleotide sequence accession numbers. The GenBank accession number of AdVVV-Ga is KC794539 and that of AdVVV-Japan is KC794540.

Acknowledgments

This work was supported by the Natural Sciences and Engineering Research Council of Canada to P.T.; H.T.P. acknowledges tuition waivers at INRS.

References

1. 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 po-tential disaster. Zootaxa. 3504:67-88.

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. 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.

4. 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.

5. Rosario K, Dayaram A, Marinov M, Ware J, Kraberger S, Stainton D, Breitbart M, Varsani A. 2012. Diverse circular ssDNA viruses discovered in dragonflies (Odonata: Epiprocta). J. Gen. Virol. 93:2668-2681.

6. Rosario K, Marinov M, Stainton D, Kraberger S, Wiltshire EJ, Collings DA, Walters M, Martin DP, Breitbart M, Varsani A. 2011. Dragonfly cyclovirus, a novel single-stranded DNA virus discovered in dragonflies (Odonata: Anisoptera). J. Gen. Virol. 92:1302–1308.

7. Johne R, Fernández-de-Luco D, Höfle U, Müller H. 2006. Genome of a novel circovirus of starlings, amplified by multiply primed rolling-circle amplification. J. Gen. Virol. 87:1189–1195.

8. Kapoor A, Dubovi EJ, Henriquez-Rivera JA, Lipkin WI. 2012. Complete genome sequence of the first canine circovirus. J. Virol. 86:7018.

9. 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.

10. Timchenko T, Katul L, Aronson M, Vega-Arreguín JC, Ramirez BC, Vetten HJ, Gronenborn B. 2006. Infectivity of nanovirus DNAs: induction of disease by cloned genome components of Faba bean necrotic yellows virus. J. Gen. Virol. 87:1735–1743.

11. Timchenko T, Katul L, Sano Y, de Kouchkovsky F, Vetten HJ, Gronen-born B. 2000. The master rep concept in nanovirus replication: identification of missing genome components and potential for natural genetic reassortment. Virology 274:189–195.

12. 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.

13. Martin DP, Lefeuvre P, Varsani A, Hoareau M, Semegni JY, Dijoux B, Vincent C, Reynaud B, Lett JM. 2011. Complex recombination patterns arising during geminivirus coinfections preserve and demarcate biologically important intra-genome interaction networks. PLoS Pathog. 7:e1002203.

14. Buchmann RC, Asad S, Wolf JN, Mohannath G, Bisaro DM. 2009. Geminivirus AL2 and L2 proteins suppress transcriptional gene silencing and cause genome-wide reductions in cytosine methylation. J. Virol. 83: 5005–5013.

15. de Vega M, Lázaro JM, Mencía M, Blanco L, Salas M. 2010. Improvement of phi29 DNA polymerase amplification performance by fusion of DNA binding motifs. Proc. Natl. Acad. Sci. U. S. A. 107:16506–16511.

16. 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.

17. Huang X, Madan A. 1999. CAP3: A DNA sequence assembly program. Genome Res. 9:868-877.

3.4.3.3. Genomic structure of volvoviruses

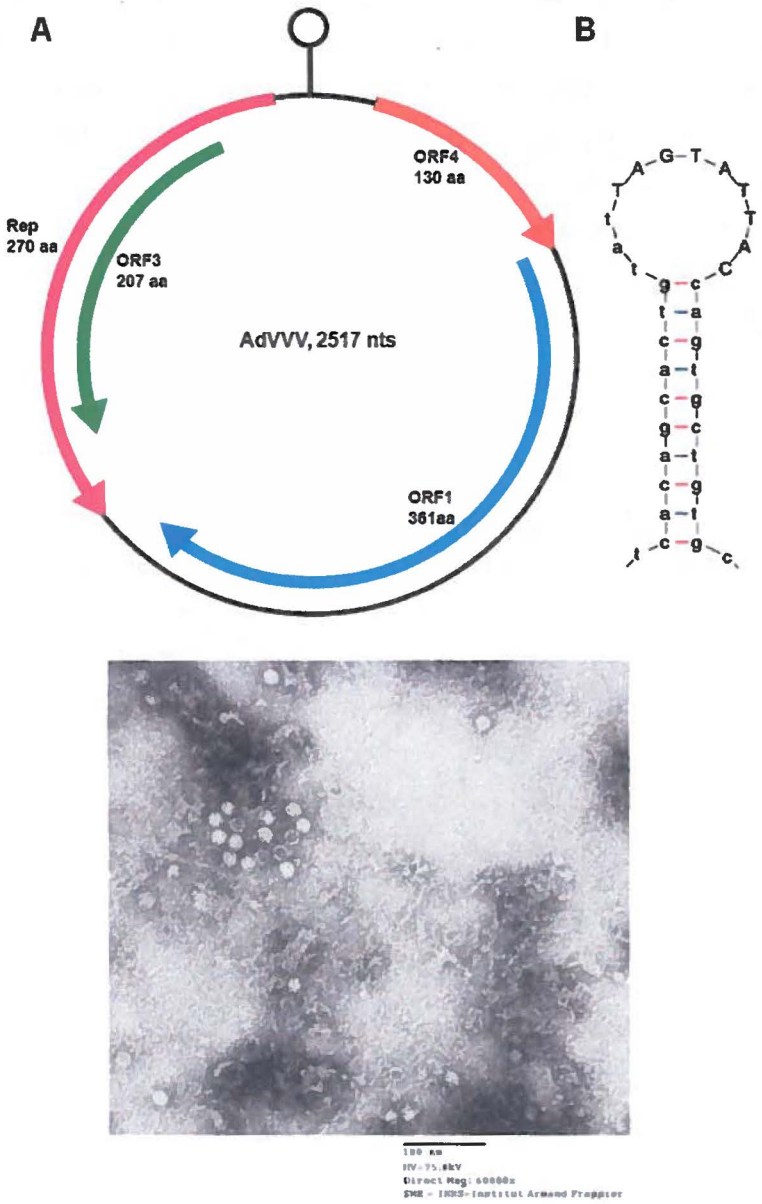


Figure 3-19. *Acheta domesticus* volvovirus – AdVVV. (A). Genome organization; (B). Typical stem-loop structure with nanonucleotide (letters in print). (C). Purified virions of AdVVV visualized under EM are about 18 nm.

3.4.4. New circo-like virus from shrimps

3.4.4.1. Publication 9: A circo-like virus isolated from *Penaeus monodon* shrimps

Genome Announc. 2014 Jan 16; 2(1)

Pham HT[1], Yu Q[1], Boisvert M[1], Van HT[2], Bergoin M[1], Tijssen P[1]

[1]INRS, Institut Armand-Frappier, Laval, QC, Canada

[2]Institute of Tropical Biology, Vietnam Academy of Science and Technology, Ho Chi Minh City, Vietnam

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.

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 reassign-ment 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 ø29 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 circularvirus (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/s00705-012-1 391-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/10.1016/0042-682 2(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.00 328-13.

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, Blanco L, Salas M. 2010. Improvement of 29 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.

3.4.4.2. Genome structure of shrimp PmCV-1 and PmaCV-2

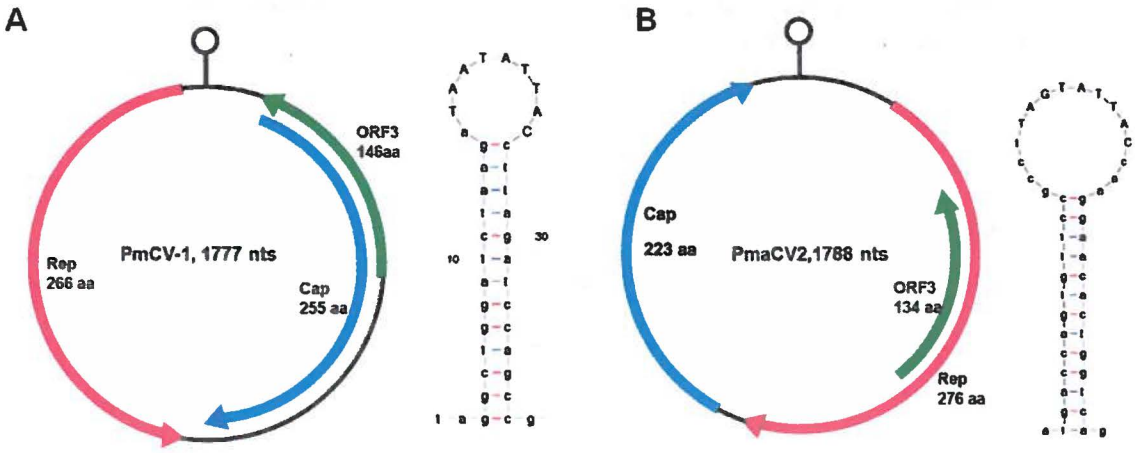


Figure 3-20. Genome organization and stem-loop structure of PmCV-1 (A) and *Peneaus monodon*-associated circovirus-2 (PmaCV-2) (B). Stem-loop structure contains the conserved nanonucleotides (letters in print) for nicking activity and the inverted repeats.

3.4.5. A small ambisense densovirus from the house cricket

3.4.5.1. Publication 10: A novel ambisense densovirus, *Acheta domesticus* mini ambidensovirus, from crickets

Genome Announc. 2013 Nov 7; 1(6)

Pham HT, Yu Q, Bergoin M, Tijssen P

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

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.

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 volvoviruses, 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 pBlueskript 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.

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 billionssold: 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* densoviruses from different outbreaks in Europe, North America, and Japan. Genome Announc. 1(4):e00629-13. doi: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.

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.

8. 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. doi: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.

3.4.5.2. Genome organization of AdMADV and comparison to members of the *Densovirus* genus

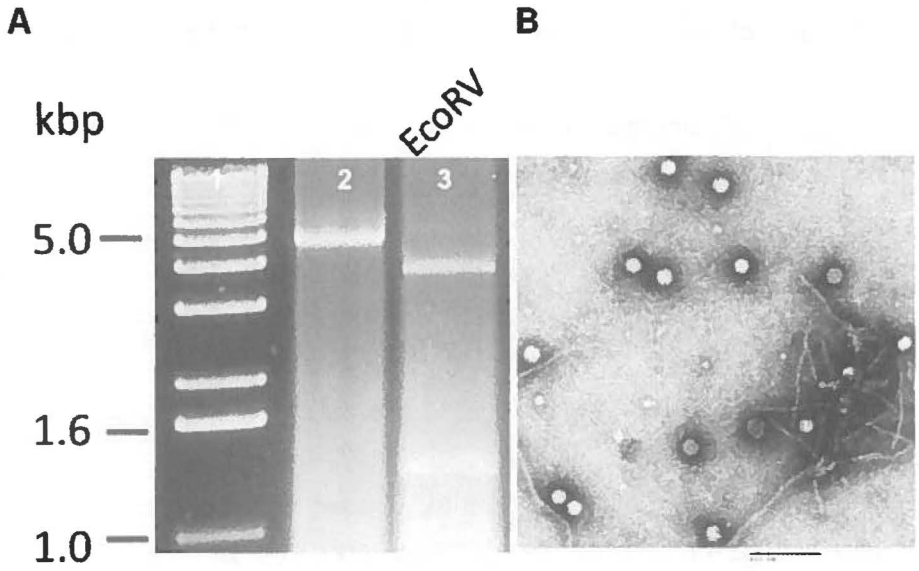


Figure 3-21. Electrophoresis analysis of AdMADV (A) and purified virions observed by EM are about 23-24 nm in diameter (B). Lane 1: 1kb ladder, lane 2: extracted DNA from purified virus is about 5 kb in size, lane 3: two distinct bands after restriction of viral DNA by EcoRV.

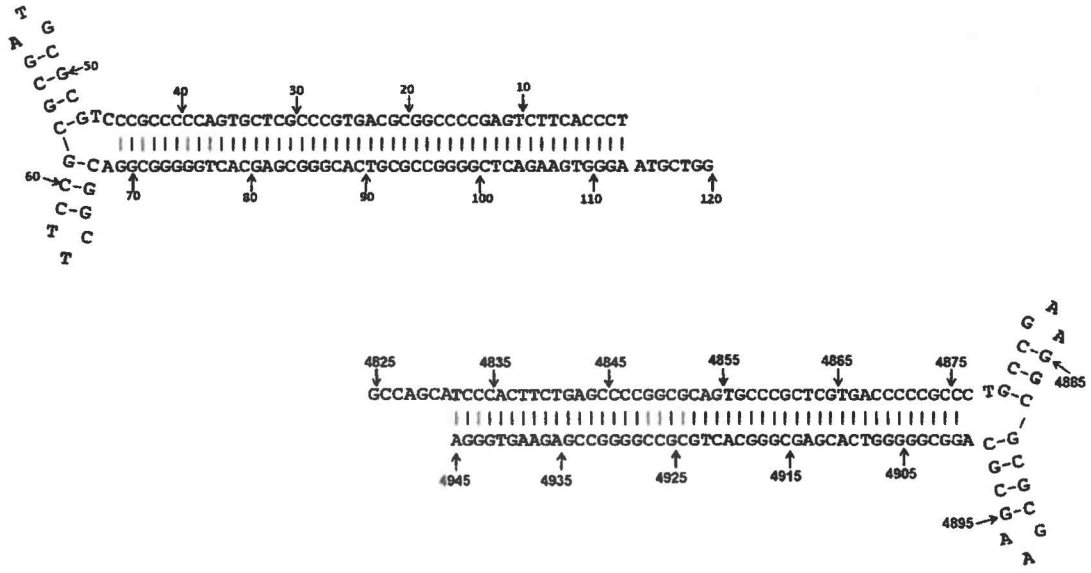
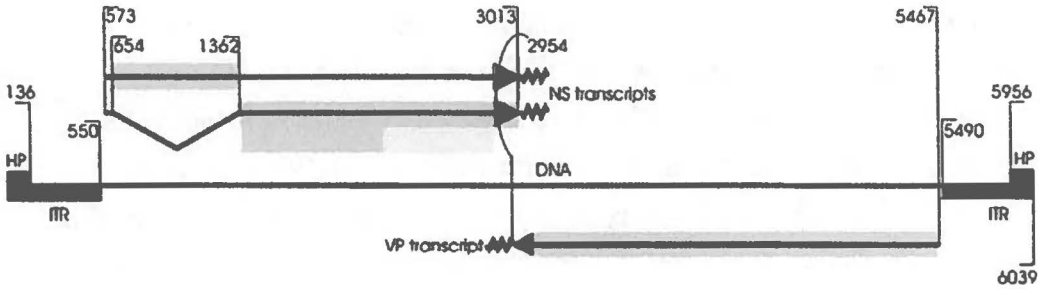


Figure 3- 22. Left and right telomeres of AdMADV (199-nt ITR) have 63% GC content and form into Y-shaped hairpin.

A. GmDNV, an ambisense densovirus



B. AdMADV, an ambisense densovirus resembling GmDNV but hardly any sequence identity

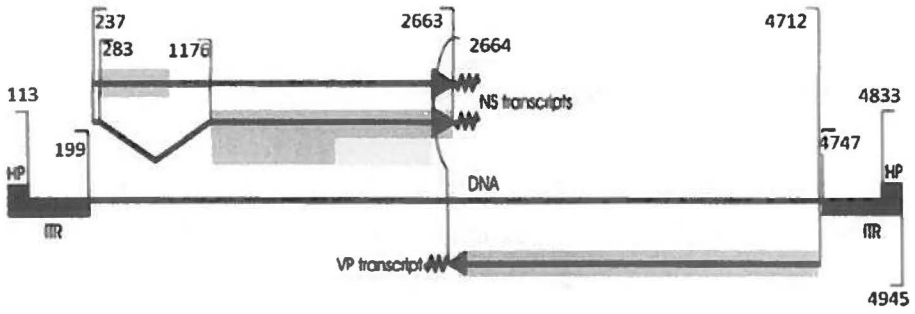


Figure 3-23. Comparison of genome organization between GmDNV (*Densovirus* genus) and AdMADV.

3.4.6. A brevidenso-like virus, *Acheta domesticus* segmented densovirus-AdSDNV, isolated from diseased crickets in US and Canada

The AdDNV epidemic appears to have diminished since the end of 2012. However, some cricket farms in US and Canada still struggle with unknown diseases. Here we describe the genome characterization of a novel brevidenso-like virus that is related to the recent epidemics occurring infrequently in cricket farms in Ontario. Since its genetic characteristics are hitherto exceptional among parvoviruses, a special taxonomic classification may be needed for this virus. We reported the method used for identification of unknown viruses previously (Pham *et al.*, 2013a, Pham *et al.*, 2013b). In short, double-stranded DNA was synthesized, followed by restriction by MboI to obtain small fragments for cloning into a BamHI-digested vector. Sequencing of clones containing inserts from both directions using universal primers and blastxn analysis to search for identity at the amino acid level confirmed the presence of a new densoviral DNA in infected crickets. Electron microscopic visualization of purified virions from these samples revealed the abundance of small particles about 22-23 nm in diameter (Figure 3-24B).

3.4.6.1. Cloning and sequencing of AdSDNV

Migration of viral DNA, extracted from different infected samples, by agarose gel electrophoresis suggested a similar structure as circular plasmid (Figure 3-24A, lane 4). However, treatment with either Phi29 DNA polymerase or T4 DNA polymerase/Large fragment of Klenow in the addition of dNTP and hexamers generated a distinct band over 3 kb (Figure 3-24A, lanes 2 and 3). Phi29 is a highly efficient DNA polymerase with circular template (Dean *et al.*, 2001, Johne *et al.*, 2009), yielding concatemeric products. However, similar amount of amplified products, with similar sizes, yielded with Phi29 polymerase or T4 DNA polymerase/Large fragment of Klenow suggested that native DNA virus is not a covalently circular form. Sequencing results from cloning of MboI-digested fragments revealed the existence of SacI in viral genomic sequence. Restriction of viral DNA by SacI showed two bands of less than 3 kb and less than 0.4 kb (data not shown). Cloning of these bands were successful in pBlueskript KS(+) (SacI and EcoRV ends). Sequencing results from both directions assembled by the CAP3 program confirmed a complete fragment of 3234 nts in length with 45% GC content. Restriction maps with HindIII, SalI also confirmed the obtained sequencing results.

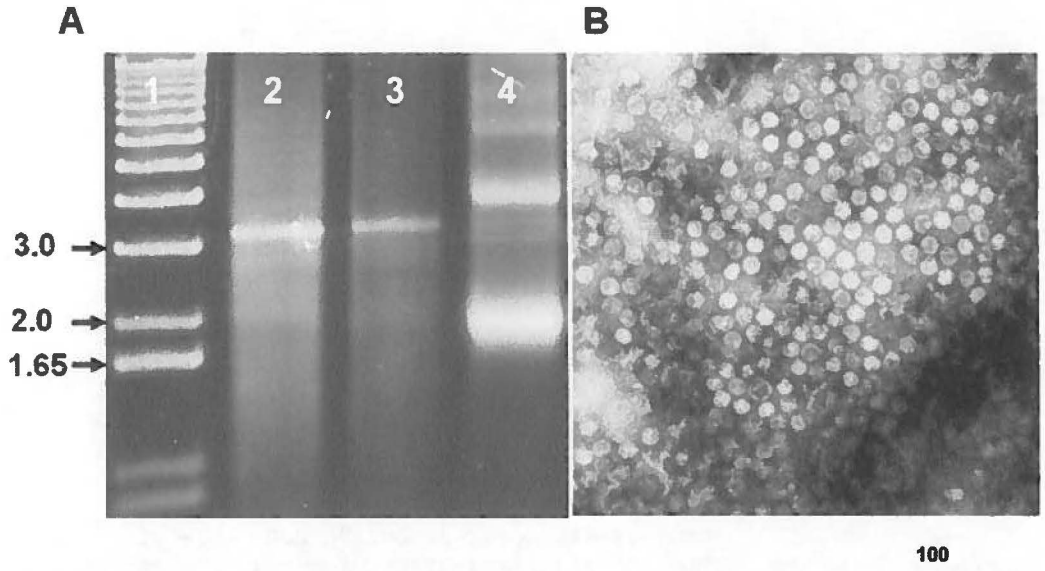


Figure 3-24. (A). Electrophoresis anlalysis of AdSDNV DNA. (B). High abundance of intact virions and empty capsid from infected crickets observed by EM. Native viral DNA (lane 4) and double stranded DNA products after amplification by ø29 (lane 2) or T4 DNA polymerase/Klenow (lane 3).

3.4.6.2. Organization of ORFs and non-coding sequences

The ORF Finder program (NCBI) showed only two major ORFs for NS1 and NS2 proteins (Figure 3-26). The largest ORF of 795 aa starts from nt 392 and terminates at nt 2779 (frame +1). This aa sequence shares an identity of 38% to NS1 proteins of CppDNV isolates (97% of query cover). At the N-terminus of this protein, from aa 88-162, common RCR motifs are located (Figure 3-25A) and from aa 582-676, Walker A, B and C of helicase motifs are found (Figure 3-25B). The second ORF was found on frame +2, starting with ATG at nt 306, 86 nts upstream of ATG of NS1 with a stop codon at nt 1615-1616. Therefore, this ORF has a coding capacity for a 436-aa protein. This aa sequence showed 40% homology to the NS2 protein of AalDNV with a covery of 95%. The NS2 start codon of AdSDNV is found upstream of NS1. This position is differrent from that of other brevidensoviruses.

Similar to other brevidensoviruses, although the sequence of the left hairpin appears to be incomplete, the telomeres at both ends are dissimilar. The right hairpin is 220 nts in size and folds into perfect, asymmetric T-shaped structure (Figure 3-27B) while on the left extremity, an incomplete hairpin of 50 nts was detected (Figure 3-27A). The left and right hairpins are situated at 235 nts downstream of NS1 stop codon and 342 nts upstream of the NS1 start codon, respectively.

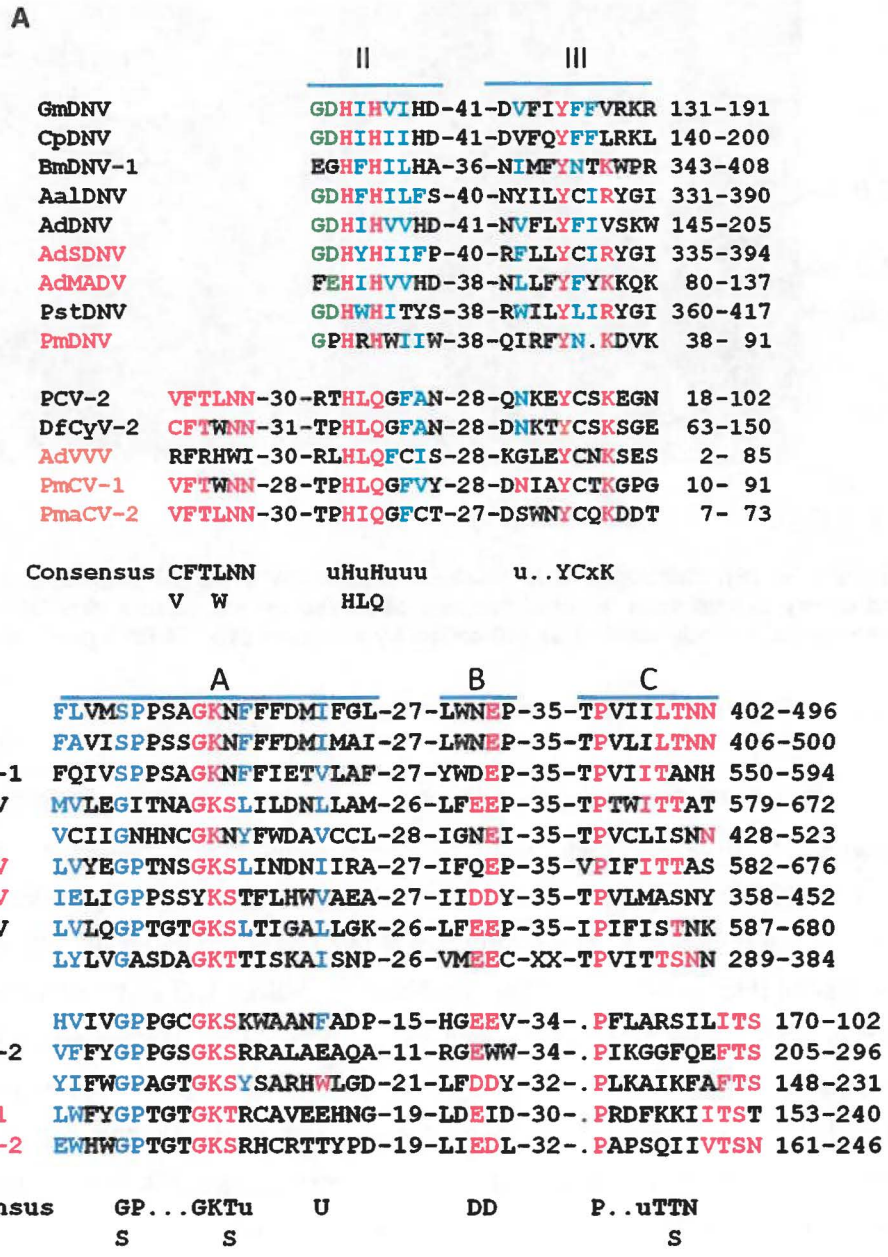


Figure 3-25. Conserved RCR (A) and Superfamily 3 helicase motifs (B) in Rep proteins of new ssDNA viruses. U = I, L, V, M, E, Y, W. Color range from red to blue indicate highly to less conserved residues.

The right telomere contains a 180 nt-palindromic sequence and the total GC-content is about 46%. The arms of right hairpin are 36 nts, 4 nts longer than those of left hairpin (32 nts). A potential promoter region was identified at nt 341-391, containing a TATA-box and Inr-box. This promoter is at a favorable position to drive the expression of both NS proteins. There is one potential poly(A) motif 2 nts downstream of the stop codon of NS1 (position 2782-2787) that may control the termination transcription of NS genes.

Two unusual characteristics made this virus unique among typical parvoviruses so far: first, the absence of a VP-ORF downstream of NS genes; and second, the NS-coding region is flanked by stable hairpin structures. Question needs to be raised here: is there another DNA segment harbouring the VP gene? Attempts to identify the VP gene of this virus are being carried out in our lab.

Figure 3-26. Genome organization of AdSDNV. Schema showed NS1 and NS2 coding region is flanked by terminal hairpins at both ends. RCR and ATPase motif are indicated on NS1 and no significant ORF encoding VP protein was found.

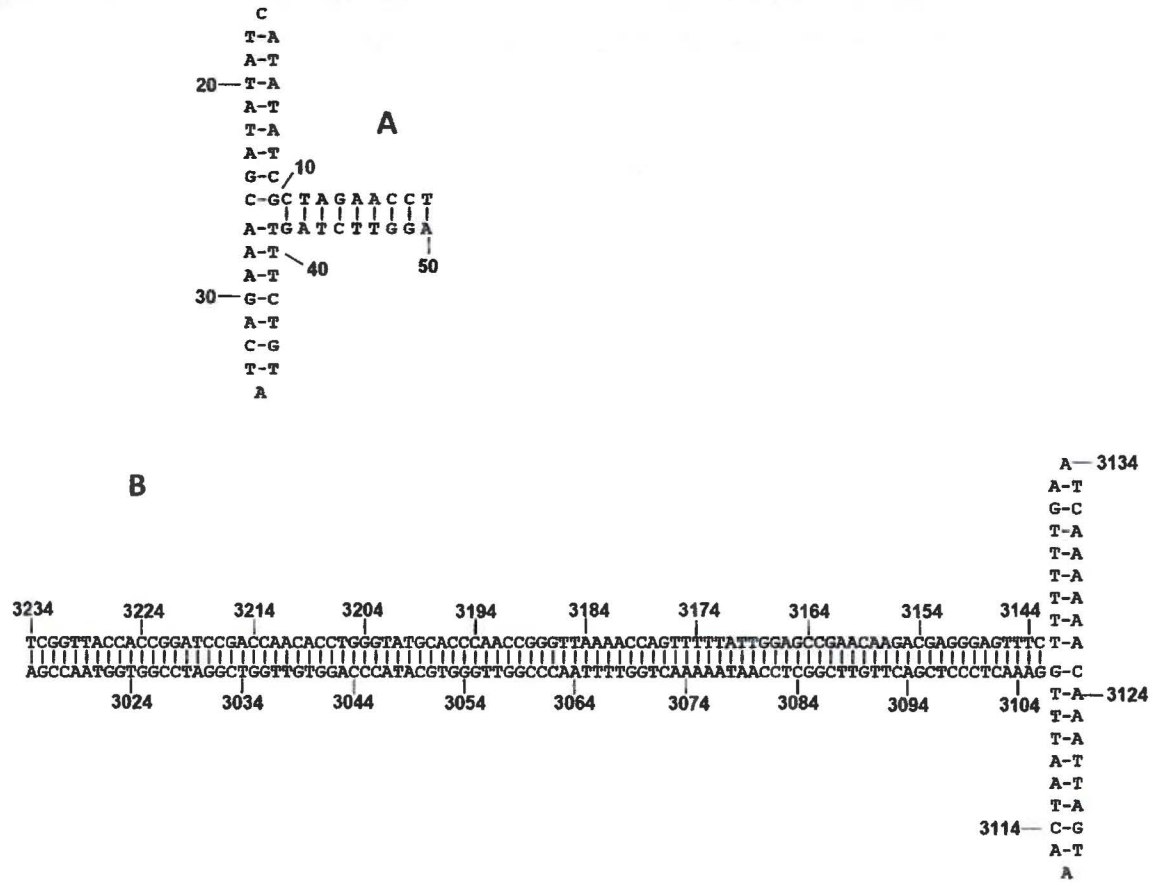


Figure 3-27. Nucleotide sequence and T-shaped structure of terminal hairpins of AdSDNV. (A) Incomplete left hairpin; (B) Complete right hairpin has a perfect T-shaped structure with 46% of GC content.

3.4.7. Relevance of our results on the discovery of new ssDNA viruses on cricket and shrimp

Using a modified sequence-independent method, we have identified three new ssDNA viruses, AdVVV, AdMADV, and AdSDNV; from AdDNV-negative crickets that have been showing symptoms of viral infection such as sluggish, anorexia, and die slowly (reported by cricket farmers). Complete viral genomes were successfully cloned and sequenced. Sequence analysis of these viruses displayed no nucleotide similarity to any viral genome on Genbank. AdVVV is the first CRESS-DNA virus reported in crickets that infects species *A. domesticus,* and *G.assimils*. This virus also has wide geographical distribution (Canada, US and Japan). Based on its genome size and ORF characteristics, AdVVV is not related to any recognized CSS-DNA viruses, and therefore, we propose a tentative name, "volvovirus" for this virus. Two other novel densoviruses discovered here, AdMADV and AdSDNV were found to have distinct characteristics compared to known DNVs. AdMADV genome is 20% smaller than GmDNV, but has similar ORF organization. AdSDNV is probably a fragmented virus since no VP gene was detected downstream of NS region, as found in normal parvoviruses. Indeed, the NS coding fragment is flanked by two hairpin structures at both ends.

Evidence of high prevalence of SS-DNA viral genomes in shrimp population has been shown with the finding of two new CRESS-DNA viruses; PmCV-1 and PmaCV-2. PmCV-1 is an ambisense virus, closely related to cycloviruses. The second CRESS-DNA virus, PmaCV-2, has a monosense genome, and is related to type IV of CRESS-DNA viruses (based on the report from Rosario, 2012).

According to our knowledge, this is the first report demonstrating the identification of circoviruses in shrimps and crickets. However, to clarify whether these viruses cause active infection or epidemy, future research must be focused on their pathogenicity and biological characteristics. The discovery of these novel viruses also implies a much wider diversity and rapid evolution of animal SS-DNA viruses than previously recognized.

4. Discussion

4.1. Comparison of transcription profiles of PstDNV and AalDNV

The transcription strategy and promoter structure of mammalian parvoviruses has been more extensively studied than that of densoviruses. Transcript maps have been fully characterized for MlDNV and GmDNV from the *Densovirus* genus (Fediere *et al.*, 2004, Tijssen *et al.*, 2003) and a few studies have been conducted for brevidensoviruses (Afanasiev *et al.*, 1994, Dhar *et al.*, 2011, Dhar *et al.*, 2010, Ward *et al.*, 2001a, Ward *et al.*, 2001b) and pefudensovirus (Mukha *et al.*, 2006, Yamagishi *et al.*, 1999, Yang *et al.*, 2008) but are lacking for other genera. So far, invertebrate parvoviruses have been shown to use mainly alternative splicing and leaky scanning to produce viral proteins (Baquerizo-Audiot *et al.*, 2009, Fediere *et al.*, 2004, Tijssen *et al.*, 2003, Yamagishi *et al.*, 1999) (see Table 4-1). Table 4-1 demonstrates the homogeneity of properties within the genera, but it also shows pronounced distinctions within the subfamily of *Densovirinae*. The known characteristics of AalDNV and PstDNV, particularly their resemblance in genome size, ORF size and ORF distribution, which were unlike that of other densoviruses, suggested a joint classification in the *Brevidensovirus* genus.

In the first part of this thesis, transcription of two brevidensoviruses, PstDNV and AalDNV, have been investigated. Moreover, the activities of their viral promoters in insect and mammalian cells were analyzed. Previously, the organization and promoter analysis of some nonviral shrimp genes using HeLa cells had been reported (Chen *et al.*, 2006). This was confirmed by luciferase assays of the putative PstDNV promoters in HeLa and LD652 cells indicating that these cell lines can also be used as a temporary surrogate cell culture systems for molecular studies of PstDNV.

A transcription map of PstDNV, and its promoters, has been recently reported (Dhar *et al.*, 2010, Dhar *et al.*, 2007). However, those results were problematic since: (i) the reported NS1 transcript termination sites were within the NS1 ORF and lacked the canonical termination motifs; (ii) the positions of poly(A) sites are in A-rich sequences (12/15 and 8/9 for NS1 and 8/11 and 9/10 for NS2) possibly leading to nonspecific priming in the 3'RACE; and since (iii) their Northern blots showed a strong band at the 4.1 kb genome size indicative of contaminating DNA. The reported poly(A) was in frame within the ORFs and would lead to the addition of long lysine tails in the translation products.

Initially, we were faced with a problem of mispriming in transcript mapping of AalDNV with incorrect position of polyadenylation site in A-rich regions (2640–AAAAAGAAAA- 2649 and 2791-AAAACACAAAAA-2802) that were prone to false priming. The A-content of the AalDNV

genome is 41%, but somewhat less for PstDNV at 36%. However, A-stretches are common in both genomes. Transfection of the plasmid that contained 3783 nts of the PstDNV sequence, but lacked the terminal hairpins of the genome, into LD652 and C6/36 cell lines allowed a precise mapping of the PstDNV transcripts by Northern blotting, RACE, amplicon mapping, and RPAs. Contrary to a previous report (Dhar *et al.*, 2010), all transcripts of PstDNV co-terminated at nt 3592. For PstDNV, this position is near the last AATAAA site at nt 3577 and appeared to be the most legitimate polyadenylation site for PstDNV-mRNA since it is located near two canonical CAYTG sequences (CATGG and CACTT at nt 3663 and 3673) and a downstream G/T-rich sequence (GGTGGG, nt 3641) (Shike *et al.*, 2000). Interestingly, the 3662-AATAAA site of AalDNV overlaps the stop codon of the VP ORF, an unusual trait shared with some other densoviruses (iteraviruses), but seems to lack the other canonical motifs. There is another potential 3651-AATAAA site in AalDNV, just 11 nts upstream of the latter.

The main difference between PstDNV and AalDNV transcription is found in the expression of the NS proteins. The PstDNV uses the P2 and P12 promoters for NS1 and NS2 whereby the P12 promoter of NS2 is mostly located within the 135-nt intron sequence of NS1. The NS1 and NS2 transcript starts were thus 365 nts apart. In contrast, AalDNV used one promoter region with closely overlapping elements to start transcription of NS1 and NS2 at the positions that are just 14 nts apart. The AalDNV NS transcripts were not spliced. In both cases, NS2 overlapped the N-terminal half of the NS1 ORF. The sequence of AaeDNV (Afanasiev *et al.*, 1991), that is closely related to AalDNV, has an NS1 ORF starting at nt 129 instead of nt 335 for AalDNV and would have a 57-aa N-terminal extension. Interestingly, at the equivalent position of nt 335 for AalDNV, an ATG (at nt 300) is present in the AaeDNV ORF and the two NS1 products from these positions would yield MESVCSEHSPCEH... (AalDNV) and MNSVCVEHSPCEH... (AaeDNV). The construction of a fusion protein of β-galactosidase inserted at nt 270 in the putative AaeDNV NS1 transcript, upstream of the corresponding NS1 transcript in AalDNV, did not give any enzyme activity (Afanasiev *et al.*, 1994). However, insertion of β-galactosidase in the NS1 frame downstream of the NS2 promoter gave a high activity suggesting that for AaeDNV, a P7/7.4 promoter would also be responsible for transcription of both NS1 and NS2 genes.

In addition to mapping of the transcripts, the activity of the PstDNV and AalDNV promoters was determined in insect and mammalian cells. The initiator (Inr) core promoter sequence motif, TCA(G/T)T(C/T) was shown to be important to the function of RNA polymerase II in arthropods in the presence or absence of TATAA sequence (Blissard *et al.*, 1992, Cherbas

et al., 1993, Pullen *et al.*, 1995, Weis *et al.*, 1992). Both Inr motifs and TATA/TATA like-boxes were present in all NS promoter regions of PstDNV and AalDNV. However, no clear TATA-like motifs sequence were found upstream of the initiator sequence CAGT of structural transcripts of AalDNV, AaeDNV or PstDNV suggesting that these promoters were under the control of downstream promoter elements (DPE). Insect genes often lack TATA-boxes and DPE functions cooperatively with the Inr for the binding of TFIID in core promoters. It has been observed that insect gene promoters frequently use the DPE as a core promoter element which is functionally analogous to the TATA- box (Burke *et al.*, 1998, Kutach *et al.*, 2000). These authors also noted that the DPE sequence motif is as common as the TATA box in *Drosophila* promoters. It is interesting that the NS1 promoter has both a conserved TATA-box and a DPE with the correct spacing, while the NS2 has a TATA-box but no conserved DPE, and the VP promoter has neither. In a database of 205 Drosophila core promoters, it was estimated that about 31% of them do not appear to contain either a TATA-box or a DPE (Smale *et al.*, 2003). RNAi depletion and overexpression experiments by Hsu *et al.* revealed a genetic circuit that controls the balance of transcription from these two core promoter motifs, the TATA-box and the DPE (Hsu *et al.*, 2008). In this circuit, they found that TATA-box binding protein (TBP) activates TATA-dependent transcription and represses DPE-dependent transcription, whereas modifier of transcription 1 (Mot1) and negative cofactor 2 (NC2) block TBP function and thus repress TATA-dependent transcription and activate DPE-dependent transcription. This regulatory circuit is probably the way by which the virus network can transfer transcriptional signals, such as those from DPE-specific and TATA-specific enhancers, via distinct pathways. There was a significantly lower level of luciferase activity under the control of P61 and P60 compared to those from P2 and P7/7.4 promoters that could be a result of the location of the reverse primers used to produce the amplicons of these promoters (lack or incomplete DPE). A previous report also noted a very high (37-fold) increase in AaeDNV VP expression when both complete viral genome termini were present in transactivation experiment (Ward *et al.*, 2001b). This could be largely due to replication of the transgene.

The expression strategy is different and may indicate that PstDNV should not be classified as a brevidensovirus although the genome size and ORFs of PstDNV resembled that of brevidensoviruses, such as AalDNV. NS1 of PstDNV and AalDNV, despite a low sequence identity, contained replication initiation motifs and DNA-dependent ATPase motifs (helicase SF3) (Fediere *et al.*, 2002), but these motifs are not restricted to parvoviruses and are found in diverse replicons from eukaryotes to eubacteria and a wide range of viruses (Gorbalenya *et al.*, 1989, Ilyina *et al.*, 1992). These two viruses do not contain the phospholipase A2 motif found in

the VP of most parvoviruses (Canaan *et al.*, 2004, Zadori *et al.*, 2001). Moreover, in our preliminary experiments, the genome of PstDNV was circularized and amplified across the ligated terminal sequences. Sequencing results of amplified products did not reveal any extra hairpins except long direct terminal repeats of 81 nts at the ends. Similar results were also obtained by the adapter ligation method that was successfully applied for ITRs of JcDNV and PiDNV. Thus, an alternative replication model using DTR was proposed in which the DTR of both plus and minus strands may anneal together to create linear or circular DNA and these forms can be served for typical RCR process. However, important experiments remain to be done in order to confirm this hypothesis.

Table 4-1. Comparison of genome and transcription characterstics between different genera of densoviruses

Genus	Genome		NS		VP	
	Size	Hairpins	Promoters	Products	Promoter	Products
Densovirus	6 kb antisense	0.55 kb ITR, T-shaped	P9	Unspliced : NS3 Spliced : NS1/NS2*	P91	Unspliced: VP1-4*
Cupidensovirus	5.75 kb antisense	0.285 kb ITR, J-shaped	P7 P17	NS3 NS1/NS2* (+ splicing)	P88	Unspliced: VP1-3*
Pefudensovirus	5.5 kb antisense	0.2 kb ITR, U-shaped	P4	Unspliced : NS3 Spliced : NS1/NS2*	P94	2 ORFs VP1-4* and splicing
Iteravirus	5 kb monosense	0.25 kb ITR, J-shaped	P7	NS1/NS2*	P58	VP1-4*
Hepandensovirus	6.3 kb monosense	0.22 kb Hairpins, no ITR	P2 P22	NS2 NS1	P48	VP1
(AalDNV) Brevidensovirus	4 kb monosense	0.15 kb, T-shape Hairpins, no ITR	This study P7/7.4	2 ORFs (790 and 363 codons)	This study P60	1 ORF (355 codons)
(PstDNV) Brevidensovirus?	4 kb? monosense	?	This study P2 and P12	2 ORFs (666 and 363 codons)	This study P61	1 ORF (329 codons)

*Indicates leaky scanning

4.2. Unique characteristics of Pefudensoviruses: expression of VP proteins from split ORFs

There appears to be at least two expression mechanisms, alternative splicing and leaky scanning that ambisense densoviruses use in order to express their proteins from a small genome (Baquerizo-Audiot *et al.*, 2009, Fediere *et al.*, 2004, Tijssen *et al.*, 2003). As mentioned previously, the transcription map of *Densovirus* members only including GmDNV, MlDNV and JcDNV have been studied in detail (Fediere *et al.*, 2004, Tijssen *et al.*, 2003, Wang *et al.*, 2013). These viruses share over 90% of nucleotide sequence identity and have similar genome organizations and therefore, their transcription profiles are very similar to each other. For AdDNV, its nucleotide sequence shares virtually no or very little homology to that of other densoviruses. However, the AdDNV genome organization was found similar to that of PfDNV, MpDNV, BgDNV, DplDNV and PcDNV with respect to those characteristics: there are two split VP-ORFs and PLA2 motif is located in the smaller ORF. Apart from that, no strong evidence for a definite classification of these viruses has been obtained.

Our study on AdDNV transcription demonstrated that AdDNV not only depends in both splicing and leaky scanning, but also on frameshifting strategies to express their different sets of VP proteins. AdDNV expressed its VP1 proteins by splicing and frameshifting which was later found to be similar for VP1 of BgDNV (Kapelinskaya *et al.*, 2011). This expression strategy so far has not been described in parvoviruses. Indeed, there is another striking featured the expression of VP proteins of AdDNV: two mutually exclusive spliced sites in VP-ORFs are used (two incompletely overlapping introns) leading to two splicing patterns that produce two different patterns of VP mRNA. As a result, the N-termini of VP1 and VP2 proteins are not related. This contrasts to all parvoviruses that commonly have similar N-terminus of VPs. Similarly, VP2 of BgDNV also has a unique N-terminus but it results from an unspliced transcript. In a transcription study of PfDNV from Yamagishi *et al.*, 1999, it was also mentioned that splicing of cDNA11 could result in the in frame linkage of two split VP ORFs and that this spliced cDNA represented about 70 % of the obtained cDNA sequences (Yamagishi *et al.*, 1999, Yang *et al.*, 2008). Apart from VP proteins, AdDNV shares several identical features with that of BgDNV as well as other members of *Densovirus* genus. First, the transcription patterns of NS proteins in AdDNV and BgDNV are very similar to those found in GmDNV, MlDNV, and JcDNV. This means that both viruses express NS1 and NS2 via splicing and leaky scanning while NS3 is expressed from the largest, unspliced transcript. Second, promoters controlling NS and VP transcription are found to be similar. In AdDNV, initiation of NS and VP mRNA are at 23 nts

downstream of the ITRs whereas these positions are at 24 and 23 nts in BgDNV, similar to those of *Densovirus* members. Therefore, most of the upstream promoter elements seem to be located within the ITR sequence. The nucleotide context of these transcription initiations in both viruses corresponds well to the Inr-box consensus sequence. Third, the 5'-UTR sequence of VP transcripts in AdDNV and BgDNV is shorter than that of NSs, only few nts compared to 30 and 32 nts (in AdDNV) and 42 and 35 nts (in BgDNV). Finally, VP and NS transcripts of AdDNV and BgDNV have a 3'-terminal overlapping region over 30 nts. This was also observed in all *Densovirus* members. Nevertheless, many putative variants of NS transcripts have also been detected in BgDNV but potential proteins encoded by these mRNA as well as their functions were not characterized.

Results from studies on transcription characteristics of AdDNV could provide more information to understand the biology of different densoviruses and of evolutionary relationships between them. A densovirus relatively closely related to AdDNV, PcDNV, may have a similar genomic structure and should be investigated in order to confirm its expression relatedness and to co-classify it with AdDNV in *Pefudensovirus* genus.

4.3. Densoviral telomeres and another alternative method to obtain these sequences

Densoviruses, except for members of *Brevidensovirus*, contain palindromic sequences at each end of their genomes. These sequences form a hairpin-like structure in "flip" or "flop" configurations since the palindromes are imperfect. The ITRs and hairpins amongst viruses of different genera vary in length; structure and GC content (Table 4-2). Among all densoviruses, members of *Densovirus* genus so far have the largest ITRs (~550 nts) and a high GC content (~60%) within their telomere structures. AdMADV, a new reported densovirus, shares similar genome organization to members of *Densovirus* genus, and has up to 63% and 75% of GC content in the ITRs and hairpin telomeres, i.e. highest amongst parvoviruses (Table 4-2). Curiously, the TATA-box of AdMADV is within in the ITRs, similar to that of *Densovirus* member also. Despite their diversity, it is clear that all hairpins have similar functions in viral replication, contain elements for transcription regulation of viral genes and have sequences specific for viral DNA encapsidation (Tijssen *et al.*, 2006a). As a result, it is important to obtain full genome containing complete ITRs sequences. However, problems arise in cloning and sequencing of infectious clones, such as nucleotide deletions in ITRs during plasmid replication in bacteria or resistance to PCR amplification due to strong secondary structures. For low content parvovirus

samples, cloning of infectious clone from native DNA is virtually impossible. Recently, techniques based on strand-displacement have been proved to be useful tools in amplifying circular DNA viruses from minute amount of extracted DNA. One example is infectious AAV, which was successfully amplified from human tissues by a modified ø29-RCA termed sequence-specific linear RCA (SSLRCA). This method is based on the property of AAV genome to exist in human tissues in a circular episomal form (Schnepp *et al.*, 2009). Unfortunately, not all parvoviruses share this feature and therefore this could not be applied. Another seamless cloning method has been developed that allowed the reconstruction of complex hairpin structure of porcine parvovirus. This method is primarily useful for mutagenesis studies since it enables cloning and joining of a set of alternative fragments that can be easily generated by PCR (Fernandes *et al.*, 2009). However, this technique is not sequence-independent; therefore, it cannot be used for unknown target sequences.

Table 4-2. Properties of densoviral telomeres

Genus		Genome size (%GC)	ITR (%GC)	Telomere (%GC)	Genbank number
Ambisense	***Densovirus***				
	GmDNV	6039 (36%)	550 (47%)	Y-shaped, 136 (61%)	NC_004286
	MlDNV	6034 (38%)	543 (49%)	Y-shaped, 126 (62%)	NC_005341
	PiDNV	5990 (38%)	540 (48%)	Y-shaped, 120 (61%)	JX645046
	JcDNV	6032 (38%)	547 (50%)	Y-shaped, 130 (61%)	KC883978
	Pefudensovirus				
	AdDNV	5425 (40%)	144 (43%)	I-shaped, 114 (42%)	HQ827781
	BgDNV	5535 (40%)	217 (59%)	I-shaped, 192 (62%)	NC_005041
	PfDNV	5455 (38%)	202 (60%)	I-shaped, 120 (60%)	AB028936
	Cupidensovirus				
	CpDNV	5759 (38%)	285 (48%)	J-shaped, 68 (49%)	NC_012685
	Unclassified				
	AdMADV	4945 (38%)	199 (63%)	Y-shaped, 113 (75%)	NC-022546
Monosense	***Iteravirus***				
	BmDNV	5076 (39%)	230 (49%)	J-shaped, 159 (52%)	AY033435
	CeDNV	5002 (38%)	230 (48%)	J-shaped, 159 (51%)	NC_004288
	PpDNV	5053 (38%)	271 (45%)	J-shaped, 175 (47%)	NC018450
	SfDNV	5012 (38%)	230 (50%)	J-shaped, 161 (53%)	NC018399
	Brevidensovirus				
	AalDNV	4176 (38%)	Absent	T-shaped, L134 (31%), R182 (42%)	X74945
	AaeDNV	3978 (38%)	Absent	T-shaped, L146 (28%), R164 (45%) DTRs, 82 (44%)	M37899
	PstDNV	3912 (43%)			KF031144
	Hepadensovirus				
	PmDNV	6321 (42%)	Absent	L136 (70%), R170 (58%)	DQ002873

Ligation of a known oligonucleotide to mRNA or to cDNA is a common strategy in the analysis of cDNA or mRNA (Chenchik *et al.*, 1996, Li *et al.*, 2006). We adapted this ligation step followed by a subsequent PCR to amplify the unknown terminal sequence of PiDNV and JcDNV. Since ITRs of these viruses contain high GC and long hairpins (Table 4-2), DMSO, betaine and 7-deaza-dGTP were added into PCR mixture as additives. DMSO and betaine are organic solvents that have been proved to enhance PCR by destabilizing secondary intramolecular structures (Chakrabarti *et al.*, 2001, Jensen *et al.*, 2010). 7-deaza-dGTP is a substitute of dGTP that acts like dATP and thus reduce the strength of hydrogen bonds. This substitute was also helpful for amplification of GC rich templates at low amounts and poor quality (Jung *et al.*, 2002). A combination of these three additives has been demonstrated to be required in order to obtain unique specific bands of GC-rich containing genes (from 67% to 79%) by PCR (Musso *et al.*, 2006). In addition, for stable replication of the recombination gene, we used Sure bacteria at low incubation temperature as previously reported in cloning of entire parvoviral genomes (Zadori *et al.*, 1995, Zhi *et al.*, 2004). In the present study, we have modified these strategies and developed a promising, alternative method that could amplify and generate stable clones of the long ITRs of both JcDNV and PiDNV that have 60% of GC in the hairpin structures. We found that this approach can be widely applicable from nanogram of starting DNA material and for unidentified, complex hairpin sequences of parvoviral termini.

4.4. Genetic variants and evolution of AdDNV

Although AdDNV has caused many epizootics over the past decades in several countries in Europe and North America, there are few reports about this virus. The first report related to the first emergences of AdDNV in Europe in 1977 (Meynardier *et al.*, 1977). Eleven years later, in 1988, Styer and Hamm observed the presence of AdDNV in a commercial cricket farm in the Southeastern USA (Styer *et al.*, 1991). Since then, AdDNV epidemics have frequently occurred in Europe, but not in North America. Until 2009 to the end of 2012, the virus became ubiquitous and severe losses have been reported in many cricket facilities throughout Canada and US. In a previous comparative study of AdDNV sequences obtained from different outbreaks, Szelei *et al.* reported that European and American isolates shared 99% of nucleotide identity to each other, the American isolates have been found to be circulating in North America for a long time before the outbreaks and that they have been diverged from the European isolates around 2006 (Szelei *et al.*, 2011).

Here we reported three new AdDNV sequences of *Acheta domesticus* from Japan, *Gryllus locorojo* from the USA and *Gryllodes sigillatus* from Canada. The AdDNV strain isolated from *Gryllus locorojo* has similar nucleotide sequence to that of the AdNA12. Comparative analysis of the sequence of AdJP12 and GsCa12 isolates revealed that they also display a high degree of sequence conservation, 99.09%, and 99.02% nucleotide identity to that of the Eu1977, respectively. Both transition and transversion substitutions were observed. These substitutions were not equally distributed over the whole genome. Most non-synonymous substitutions were found in NS3, then VP, and finally NS1/NS2 genes. NS1 encodes for the largest protein (576 aa) but it is the most conserved. There were only 9 and 10 nucleotide substitutions of which 4 and 3 were non- synonymous in NS1 of AdJP12 and GsCa12, respectively. In both AdJP12 and GsCa12, the 213-aa NS3 protein included 11 nucleotide substitutions, eight of them were non-synonymous. Thus, the non-synonymous/synonymous substitutions ratio for this gene is highest, at about 2.67. Compared to NS genes, there are 26 (AdJP12) and 27 (GsCa12) substitutions on VP genes of which only 10 and 13 are non-synonymous. Therefore, in both isolates, nonsynonymous/synonymous substitutions ratio for VP genes is less than 1. These results are also in accordance with previous observations reported by Szeilei *et al*, 2011. In addition, the three conserved regions A, B (of NS1) and C (of VP) are stable in AdJP12 and in GsCa12; only one synonymous substitution was found in the C region.

Another issue is that a greatest number of non-synonymous substitutions found within VP genes localized in EF (residues 138 to 233 of VP4) and GH (residues 249 to 342 of VP4), the two largest loops of the capsid (Table 4-3) that are exposed on the capsid surface (Meng *et al.*, 2013) (Figure 4-1). GH loop accounts for the elevated region around the 3-fold spike of the parvovirus capsid. It has been shown that in both vertebrate and invertebrate parvoviruses, except for PstDNV, the GH loop plays an important role in the assembly of trimers during cytoplasmic folding for successful nuclear transportation (Lombardo *et al.*, 2000, Riolobos *et al.*, 2006, Simpson *et al.*, 1998). For CPV and FPV, the difference in transferrin receptor (TfR) binding efficiency was determined by a two aa residues (VP2-93 and -323) adjacent to the 3-fold spike region of the capsid surface and determined host range tropism of these viruses (Parrish *et al.*, 2005). In JcDNV, it was reported that 4 residues in the EF loop could control the tissue tropism of virus (Multeau *et al.*, 2012). As discussed above, the distribution of non-synonymous substitutions was observed mainly in NS3 and in certain regions on the capsid genes suggesting that the AdDNV populations have not only undergone random genetic drift but also strong selective pressures.

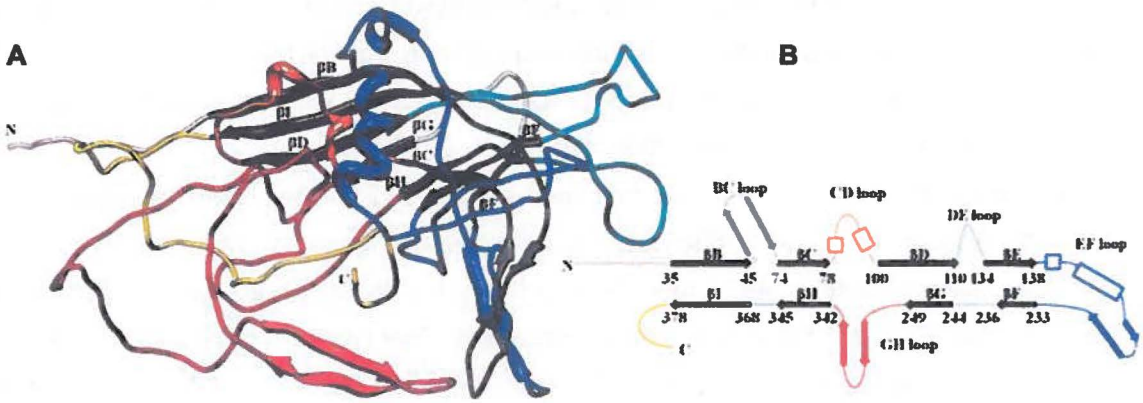


Figure 4-1. Three-dimensional structure of AdDNV capsid protein (A) and diagram of capsid protein structure (B). The core "jelly roll" is made by strands is in black. The surface loops connecting the strands are colored: BC loop, gray; CD loop, orange; DE loop, sky blue; EF loop, dark blue; and GH loop, red (Meng *et al.*, 2013).

Table 4-3. Description of amino acid residues in VP1 of eight AdDNV strains which respects to those of AdEu1997

Position	AdEu1977	AdEu2004	AdEu2006	AdEu2007	AdEu2009	NA2009/ GIUS2012	AdJP2012	GsCa2012
3	Gly	Asp*	Asp*	Asp*	Ala*			
63	Gly						Glu*	Glu*
84	Glu	Gln	Gln	Gln	Gln	Gln		
91	Thr						Ser	Ser
152	Asp	Gly*	Gly*	Gly*				
206	Thr				Ser			
249	Val	Leu	Leu	Leu	Leu	Leu	Leu	Leu
337	Val	Ile	Ile	Ile	Ile	Ile	Ile	Ile
463	Gly	Ala*			Ala*			
477	Asn		Ser					
535	Ile					Val	Val	
544	Ala	Ser	Ser	Ser	Ser			Ser
551	Ser			Thr	Ala			Thr
560	Arg							Pro*
563	Asn							Lys*
597	Phe	Leu*						
598	Ala	Pro*						
620	Asn	His	His	His	His			
665	Ser	Arg						
670	Thr	Ala*	Ala*	Ala*	Ala*	Ala*	Ala*	Ala*
679	Asp		Asn					
686	Ile				Val			
693	Lys	T*	Thr*	Thr*	Thr*	Thr*	Thr*	Thr*
695	Asn			Lys*	695Lys*	Lys*	Lys*	Lys*
727	Lys	Arg	Arg	Arg	Arg	Arg	Arg	Arg
798	Thr							Ala*
800	Glu			Gln		Glu	Gln	

*Non-similar substitutions

AdJP12 and GsCa12 have two non-synonymous substitutions 63Glu and 91Ser at the N-terminal of VP1, which were not found in other AdDNV strains. The capsid of GsCa12 also contained three unique non-synonymous mutations 560Pro, 563Lys and 798Ala (VP1 residues) that are not found in AdJP12 and other isolates (Table 4-3). Comparing the NS genes of AdJP12 and GsCa12, we observed only 5 substitutions downstream in the NS1/NS2 sequence whereas the NS3 sequences were completely identical. This also reflects the close relationship between these strains and the presence of a separate clade of AdJp12 and GsCa12 from the others on the phylogenetic tree.

It is also important to note that, previous AdDNV strains were only isolated from house cricket from Europe and North America. The finding of new isolates from Japanese house crickets and from *Gryllodes sigillatus* in Canada extends the earlier study of sequence analysis and phylogeny of AdDNV isolates. It clearly indicated that AdDNV is a cosmopolitan virus that can infect other cricket species rather than only *Acheta domesticus*. Recently, in the commercial market of crickets, *Gryllus assimilis* has been replaced by *Gryllodes sigillatus*, which is claimed to be resistant to AdDNV (http://www.ghann.com/new_crickets.cfm). However, caution is warranted to prevent a possible emergence of future epidemics by GsCa12. As seen for catastrophes caused by AdDNV in house cricket, this virus needs to have a long co-existence time before establishing stable interactions with cricket host. Moreover, as mentioned above, AdDNV can be commonly transmitted by different routes i.e. fecal-oral-aerosol and therefore, the presence of virus in crickets as well as other co-cultured insects and materials from rearing facilities should be verified to prevent significant virus transmission and minimize economic impact.

4.5. Diversity of ssDNA viruses in crickets and shrimps

Population density significantly influences the virus infection in host populations. Obviously, the monoculture of crickets and shrimps on a large-scale in rearing farms has induced the possible potential of viruses to establish stable, wide-scale infections, rapid accumulation of virus in the environment and rapid transmission to all members of the host population. Moreover, the poor health status of cultured crickets and shrimps due to the high population pressure can lead to co-infection of multiple viruses, giving the opportunity for genetic exchange as recombination and mutation do occur. In fact, these processes are

common place and can create new pathogenic strains or new virulent viruses that can infect new hosts (Kautsky *et al.*, 2000, Parrish *et al.*, 2008, Shackelton *et al.*, 2007).

Several methods have been developed to discover unrecognized viruses based on traditional techniques including sequence-independent single primer amplification (SISPA), cDNA-amplified fragment length polymorphism (cDNA-AFLP) (Allander *et al.*, 2001, Pyrc *et al.*, 2008) or whole genome amplification (WGA) (Rosario *et al.*, 2012b) to high-throughput sequencing (Bellehumeur *et al.*, 2013, Yozwiak *et al.*, 2012). Traditional methods are laborious and time-consuming whereas deep sequencing is a powerful technique but it can cost thousand dollars for each run. In our present studies, we reported the detection of three new densoviruses and three CRESS-DNA viruses in both shrimps and crickets by a modified SISPA method that was successfully applied in the detection of animal bocaviruses (Allander *et al.*, 2001). Our current method includes all steps of SISPA except for adapter ligation and PCR. Instead of digestion with Csp6I, we used MboI that creates a BamHI compatible site. Several clear bands after MboI-digestion could be visualized on agarose gel and then cloned in a BamHI-vector with comparative ease.

From PstDNV-infected shrimp samples collected from Vietnam, we have identified at least two unknown CRESS-DNA viruses and one new densovirus. The complete genomes of two CRESS-DNA viruses were cloned and sequenced. For the new densovirus, we obtained the complete coding sequence except for the termini since the sequencing of these structures failed due to very high GC ratio (probably more than 70%). Therefore, data on this densovirus are not included in this thesis and more efforts will be aimed in obtaining the entire sequence of this virus. The CRESS-DNA viruses named PmCV-1 and PmaCV-2 were the first circoviruses found in shrimps. PmCV-1 is related to cycloviruses that have been discovered in dragonflies, tissues of 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. There has been some evidence of cyclovirus transmission between different host species (Li *et al.*, 2010, Li *et al.*, 2011). The second circovirus, PmaCV-2, was isolated from the same sample. According to its genome type, this virus may belong to type IV, as proposed in the report from Rosario, 2012. However, some striking aspects of this virus include the usage of a different genetic code and potential promoters of prokaryotes. If that is the case, it means that ciliates or other protozoa may be ancient or recent hosts of this virus. Since those viruses were isolated from PstDNV-positive shrimps, we cannot conclude whether they are shrimp pathogens or not. Recently, in June

2013, a report of FAO mentioned the early mortality syndrome (EMS) or acute hepatopancreatic necrosis syndrome (AHNS) related to high mortality in cultured *P. monodon* that has been happening in Asia since early 2012. Shrimp experts have been directing attention to identify the etiology of this newly emerging disease. For these reasons, origin, biological and molecular characteristics of these unusual viruses need to be further investigated.

Similar to shrimps, crickets have been found to be sensitive hosts for many viruses. Before 2012, it was very common that cricket samples from rearing farmings developed AdDNV infections. However, later, there have been more negative-AdDNV samples, even in *A. domesticus*, the most sensitive host of this virus. The first viral agent isolated from those samples was AdVVV. This is also the first CRESS-DNA virus reported in house crickets that related to small outbreaks in Canada. Shortly after, two other isolates of AdVVV have been found in *A. domesticus* from Japan and *G. assimilis* from the USA. These crickets were reported to have been shown similar visible symptoms of viral infection such as anorexia, sluggishness, and slow death.

As mentioned on pulication 8, further comparative analysis between their nucleotide sequences of AdVVVs revealed that all substitutions were in the Cap gene except only one in the intergenic region. AdVVV-Japan differs from AdVVV-Canada in 22 substitutions but 18 of them were within the Cap gene. The potential Cap protein of AdVVV was found to contain an arginine-rich region at the N-terminus. This region is conserved in all strains, especially in the Japanese one; with a total of 14 non-synonymous substitutions over the capsid protein but none of them in the N-terminus. The arginine-rich N-terminal end is a usual property of capsids in circoviruses and in many icosahedral RNA viruses such as *Bromovirus*, *Cucumovirus*, *Sobemovirus,Tombusvirus*, and in Tat & Rev, RNA-binding proteins of human immunodeficiency virus (Tan *et al.*, 1995). In circoviruses, this region encompasses nuclear targeting signals responsible for the delivery of viral DNA molecules and Rep protein to the nucleus (Heath *et al.*, 2006, Liu *et al.*, 2001). Possibly, similar functions may be involved in this region of AdVVV capsid.

The complete genome of AdMADV, a novel densovirus isolated from AdDNV-free crickets, was also cloned and sequenced. This virus has a small genome size of 4945 nts but its genome organization and ITR structure highly resembles that of HaDNV, JcDNV, MlDNV, and GmDNV, members of *Densovirus* genus. Whilst the nucleotide sequences of *Densovirus* members are highly identical, over 85%, AdMADV has no nucleotide sequence identity on Genbank. The aa level, NS1 and VP proteins share about 25% to that of *Iteravirus* members

and *Densovirus* members, respectively. However, phylogenetic analysis of NS1 protein revealed that AdMADV clustered into the same group with AdDNV and differed from other densoviruses (Figure 4-2). This may suggest a common ancestor between these two viruses and they possibly have co-existed with *A. domesticus* host. Since virus must replicate in a host in order to survive and genetic changes during evolution like mutations and recombination could expand the host range of many viruses, AdMADV could have infected house crickets as a persistent infection or other insects were its real host. It was also noted that substitution rate in parvoviruses, especially in capsid genes is as high as that in RNA viruses, around 10^{-4}/site/year (Liu *et al.*, 2011, Shackelton *et al.*, 2005, Ren et al., 2013) and recombination frequencies is also high (Lefeuvre *et al.*, 2009). In addition, though it is expected a similar expression profile with members of densovirus, AdMADV merits further studies for its classification into *Densovirus* genus or not.

Until now, all members of *Brevidensovirus* were isolated from mosquitoes, except for PstDNV. Here we report an unusual brevidensovirus from crickets named AdSDNV. We could only obtain a NS1 coding region flanked by two hairpin structures. Results after sequencing and sequence assembly confirmed the observation of a unique viral DNA band on agarose gels. It seems that AdSDNV is a bipartite single-stranded DNA genome with different VP and NS coding segments. BmDNV-2, also with bipartite genomes, is not a parvovirus since two large VD1 and VD2 fragments separately code for different types of NS (including DNA polymerase) and VP proteins and the absence of hairpins (Hayakawa *et al.*, 2000). Similar to AdMADV, AdSDNV lacks a nucleotide sequence from GenBank database. NS1 is a highly conserved protein but in this case has a low sequence identity to mosquito brevidensoviruses. Hence, AdSDNV was clustered into a subclade with AalDNV in the phylogenetic tree of NS1 proteins (Figure 4-2). Although we could later detect the presence of this virus in several cricket samples later by PCR and EM, we need more convincible data about its VP genes.

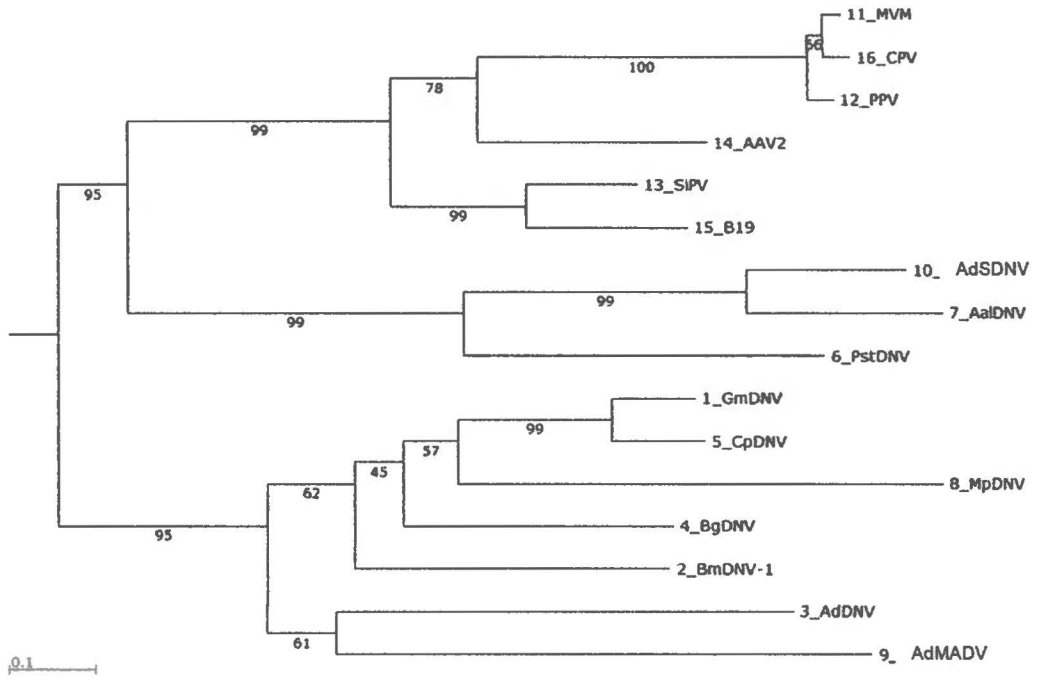


Figure 4-2. Phylogenetic tree of NS1 protein of AdMADV, AdSDNV and some typical members of *Parvoviridae* family. AdSDNV is clustered with AalDNVinto the same clade. AdMADV is distinct from other groups but may have evolved from a common ancestor with AdDNV. Bootstrapped confidence values are indicated on each branch (1000 bootstrap replicates). Scale bar indicates the rate of amino acid substitutions.

5. Conclusion

- Given the description of complete expression maps of PstDNV, AalDNV and AdDNV, this extended our insights into basic mechanisms at a molecular level of how densoviruses govern the expression of their genes for infection. These mechanisms were found to be much more complex than previously thought. Our results revealed the fact that PstDNV and AalDNV, which were earlier both regarded as members of the Brevidensovirus genus, have evolved significantly different expression strategies. Moreover, the presence of DTRs at terminal sequence of PstDNV has been not observed in parvoviruses so far. Hence, it is plausible that a new genus should be proposed for PstDNV. Likewise, since AdDNV has unique features differing from other parvoviruses; split VP ORFs and an unusual transcription strategy, translational frameshifting combined with alternative splicing to express the VP1, *Pefudensovirus* genus was proposed for AdDNV. So, apart from phylogenetic analysis, transcription profiles should be considered as a criterion for classification of parvovirus.

- Our study showed that different AdDNV outbreaks in house crickets in US, Europe and Japan caused by different strains suggested a worldwide distribution of this virus. Moreover, the recent identification of a GsCa12 AdDNV isolate in infected *G. sigillatus* indicates that this virus can infect a wide host range of cricket species.

- PstDNV and AdDNV have been well-known viruses that have caused severe impact on cultured shrimps and crickets. Here we reported the discovery and genome characterization of numerous novel densoviruses and, for the first time, circoviruses in cultured crickets and shrimps. These results strongly imply a great diversity of ssDNA-viruses infecting these economically important animals. Future work will be focused on further biological studies on pathogenicity and molecular features of theses viruses for further insight into the evolution and classification of ssDNA viruses. This will also provide more information for virus diagnosis and for the prevention and control of future emerging virus diseases.

- Finally, our methods developed for virus detection and cloning of densoviral telomeres permit the efficient rescue of viral DNA sequences. As these methods were successfully applied for infected shrimp samples that often contain limited amount of viruses, we also would suggest these methods as alternative techniques for ssDNA virus diagnosis and identification.

References

Abd-Alla A (2003) *Recherches sur un Parvovirus d'insecte, le densovirus du lépidoptère Junonia coenia (JcDNV): Étude de la cinétique du cycle viral, du rôle de la protéin NS-3 dans la réplication et du rôle des protéines de capside dans la production de virions infectieux*. Thèse du Doctorat (Université Montpellier II, France).

Abd-Alla A, Jousset FX, Li Y, Fediere G, Cousserans F & Bergoin M (2004) NS-3 protein of the *Junonia coenia* densovirus is essential for viral DNA replication in an Ld 652 cell line and Spodoptera littoralis larvae. *J Virol* 78(2):790-797.

Abraham AD, Bencharki B, Torok V, Katul L, Varrelmann M & Josef Vetten H (2010) Two distinct nanovirus species infecting faba bean in Morocco. *Arch Virol* 155(1):37-46.

Adair BM (2000) Immunopathogenesis of chicken anemia virus infection. *Dev Comp Immunol* 24(2-3):247-255.

Afanasiev BN, Galyov EE, Buchatsky LP & Kozlov YV (1991) Nucleotide sequence and genomic organization of Aedes densonucleosis virus. *Virology* 185(1):323-336.

Afanasiev BN, Kozlov YV, Carlson JO & Beaty BJ (1994) Densovirus of Aedes aegypti as an expression vector in mosquito cells. *Exp Parasitol* 79(3):322-339.

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(11):1369-1381.

Albaladejo JD, Tapay LM, Migo VP, Alfafara CG, Somga JR, Mayo SL, Miranda RC, Natividad K, Magbanua FO, Itami T, Matsumura M, Nadala ECBJ & Loh PC (1998) Screening for Shrimp Viruses in the Philippines *In Flegel TW (ed) Advances in shrimp biotechnology*.

Allan G, Meehan B, Todd D, Kennedy S, McNeilly F, Ellis J, Clark EG, Harding J, Espuna E, Botner A & Charreyre C (1998) Novel porcine circoviruses from pigs with wasting disease syndromes. *Vet Rec* 142(17):467-468.

Allan GM, McNeilly F, Cassidy JP, Reilly GA, Adair B, Ellis WA & McNulty MS (1995) Pathogenesis of porcine circovirus; experimental infections of colostrum deprived piglets and examination of pig foetal material. *Vet Microbiol* 44(1):49-64.

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(20):11609-11614.

Andersen MT, Richardson KA, Harbison SA & Morris BA (1988) Nucleotide sequence of the geminivirus chloris striate mosaic virus. *Virology* 164(2):443-449.

Anderson IG, Law AT, Shariff M & Nash G (1990) A parvo-like virus in the giant freshwater prawn, Macrobrachium rosenbergii. *Journal of Invertebrate Pathology* 55(3):447-449.

Baldauf AQ, Willwand K, Mumtsidu E, Nuesch JP & Rommelaere J (1997) Specific initiation of replication at the right-end telomere of the closed species of minute virus of mice replicative-form DNA. *J Virol* 71(2):971-980.

Bando H, Choi H, Ito Y & Kawase S (1990) Terminal structure of a Densovirus implies a hairpin transfer replication which is similar to the model for AAV. *Virology* 179(1):57-63.

Bando H, Kusuda J, Gojobori T, Maruyama T & Kawase S (1987) Organization and nucleotide sequence of a densovirus genome imply a host-dependent evolution of the parvoviruses. *J Virol* 61(2):553-560.

Baquerizo-Audiot E, Abd-Alla A, Jousset FX, 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. *J Virol* 83(13):6863-6873.

Bashir T, Horlein R, Rommelaere J & Willwand K (2000) Cyclin A activates the DNA polymerase delta -dependent elongation machinery in vitro: A parvovirus DNA replication model. *Proc Natl Acad Sci U S A* 97(10):5522-5527.

Becerra SP, Rose JA, Hardy M, Baroudy BM & Anderson CW (1985) Direct mapping of adeno-associated virus capsid proteins B and C: a possible ACG initiation codon. *Proc Natl Acad Sci U S A* 82(23):7919-7923.

Bell KE, Dale JL, Ha CV, Vu MT & Revill PA (2002) Characterisation of Rep-encoding components associated with banana bunchy top nanovirus in Vietnam. *Arch Virol* 147(4):695-707.

Bellehumeur C, Boyle B, Mandeville I & Gagnon CA (2013) High-throughput sequencing revealed the presence of an unforeseen parvovirus species in Canadian swine: the porcine partetravirus. *Can Vet J* 54(8):787-789.

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.

Biagini P (2004) Human circoviruses. *Vet Microbiol* 98(2):95-101.

Biagini P, Bedarida S, Touinssi M, Galicher V & de Micco P (2013) Human gyrovirus in healthy blood donors, France. *Emerg Infect Dis* 19(6):1014-1015.

Blinkova O, Victoria J, Li Y, Keele BF, Sanz C, Ndjango JB, Peeters M, Travis D, Lonsdorf EV, Wilson ML, Pusey AE, Hahn BH & Delwart EL (2010) Novel circular DNA viruses in stool samples of wild-living chimpanzees. *J Gen Virol* 91(Pt 1):74-86.

Blissard GW, Kogan PH, Wei R & Rohrmann GF (1992) A synthetic early promoter from a baculovirus: roles of the TATA box and conserved start site CAGT sequence in basal levels of transcription. *Virology* 190(2):783-793.

Bodendorf U, Cziepluch C, Jauniaux JC, Rommelaere J & Salome N (1999) Nuclear export factor CRM1 interacts with nonstructural proteins NS2 from parvovirus minute virus of mice. *J Virol* 73(9):7769-7779.

Bonami JR, Mari J, Poulos BT & Lightner DV (1995) Characterization of hepatopancreatic parvo-like virus, a second unusual parvovirus pathogenic for penaeid shrimps. *J Gen Virol* 76 (Pt 4):813-817.

Bonami JR, Trumper B, Mari J, Brehelin M & Lightner DV (1990) Purification and characterization of the infectious hypodermal and haematopoietic necrosis virus of penaeid shrimps. *J Gen Virol* 71 (Pt 11):2657-2664.

Boublik Y, Jousset FX & 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.

Brock JA & Lightner DV (1990) *Diseases caused by microorganisms (Diseases of Crustacea).*. Biologische Anstalt Helgoland,

Brockhaus K, Plaza S, Pintel DJ, Rommelaere J & Salome N (1996) Nonstructural proteins NS2 of minute virus of mice associate in vivo with 14-3-3 protein family members. *J Virol* 70(11):7527-7534.

Burke TW, Willy PJ, Kutach AK, Butler JE & Kadonaga JT (1998) The DPE, a conserved downstream core promoter element that is functionally analogous to the TATA box. *Cold Spring Harb Symp Quant Biol* 63:75-82.

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(15):14502-14508.

Cardona CJ, Oswald WB & Schat KA (2000) Distribution of chicken anaemia virus in the reproductive tissues of specific-pathogen-free chickens. *J Gen Virol* 81(Pt 8):2067-2075.

Caspar DL & Klug A (1962) Physical principles in the construction of regular viruses. *Cold Spring Harb Symp Quant Biol* 27:1-24.

Catap ES & Travina RD (Édit) (2005) *Experimental transmission of Hepatopancreatic Parvovirus (HPV) infection in Penaeus monodon postlarvae.* Fish Health Section, Asian Fisheries Society, Manila. 415-420 p.

Cater JE & Pintel DJ (1992) The small non-structural protein NS2 of the autonomous parvovirus minute virus of mice is required for virus growth in murine cells. *J Gen Virol* 73 (Pt 7):1839-1843.

Chaiyakul M, Hsu K, Dardari R, Marshall F & Czub M (2010) Cytotoxicity of ORF3 proteins from a nonpathogenic and a pathogenic porcine circovirus. *J Virol* 84(21):11440-11447.

Chakrabarti R & Schutt CE (2001) The enhancement of PCR amplification by low molecular-weight sulfones. *Gene* 274(1-2):293-298.

Chapman MS & Agbandje M (2006) Atomic structure of viral particles. *Parvoviruses*, (Ed.) IMLD (Édit.) Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. p 107-121.

Chen JY, Chen JC, Lin SH, Pan CY & Kuo CM (2006) Organization and promoter analysis of a tiger shrimp *Penaeus monodon* single WAP domain-containing protein gene. *Fisheries Science* 72(5):1086-1095.

Chenchik A, Diachenko L, Moqadam F, Tarabykin V, Lukyanov S & Siebert PD (1996) Full-length cDNA cloning and determination of mRNA 5' and 3' ends by amplification of adaptor-ligated cDNA. *Biotechniques* 21(3):526-534.

Cheng JH, Sheu SC, Lien YY, Lee MS, Chen HJ & Su WH (2012) Identification of the NLS and NES motifs of VP2 from chicken anemia virus and the interaction of VP2 with mini-chromosome maintenance protein 3. *BMC Vet Res* 8:15.

Cherbas L & Cherbas P (1993) The arthropod initiator: the capsite consensus plays an important role in transcription. *Insect Biochem Mol Biol* 23(1):81-90.

Cheung AK (2003a) Comparative analysis of the transcriptional patterns of pathogenic and nonpathogenic porcine circoviruses. *Virology* 310(1):41-49.

Cheung AK (2003b) The essential and nonessential transcription units for viral protein synthesis and DNA replication of porcine circovirus type 2. *Virology* 313(2):452-459.

Cheung AK (2003c) Transcriptional analysis of porcine circovirus type 2. *Virology* 305(1):168-180.

Cheung AK (2004a) Detection of template strand switching during initiation and termination of DNA replication of porcine circovirus. *J Virol* 78(8):4268-4277.

Cheung AK (2004b) Palindrome regeneration by template strand-switching mechanism at the origin of DNA replication of porcine circovirus via the rolling-circle melting-pot replication model. *J Virol* 78(17):9016-9029.

Cheung AK (2006) Rolling-circle replication of an animal circovirus genome in a theta-replicating bacterial plasmid in Escherichia coli. *J Virol* 80(17):8686-8694.

Cheung AK (2007) A stem-loop structure, sequence non-specific, at the origin of DNA replication of porcine circovirus is essential for termination but not for initiation of rolling-circle DNA replication. *Virology* 363(1):229-235.

Cheung AK (2012) Porcine circovirus: transcription and DNA replication. *Virus Res* 164(1-2):46-53.

Chia YH, Po CC, Li CC, Chien T, Maw SC, Kwo CH, Hui FK, Ming CT & Shinn ST (2006) Infectious hypodermal and haematopoietic necrosis virus (IHHNV) infections in giant freshwater prawn, Macrobrachium rosenbergii. *Aquaculture* 258(1-4):73–79.

Choi C & Chae C (1999) In-situ Hybridization for the Detection of Porcine Circovirus in Pigs with Postweaning Multisystemic Wasting Syndrome. *Journal of Comparative Pathology* 121(3):265-270.

Chong YC & Loh H (1984) Hepatopancreas chlamydial and parvoviral infections of farmed marine prawns in Singapore. *Singapore Vet. J.* 9:51-56.

Christensen J, Cotmore SF & Tattersall P (1995a) Minute virus of mice transcriptional activator protein NS1 binds directly to the transactivation region of the viral P38 promoter in a strictly ATP-dependent manner. *J Virol* 69(9):5422-5430.

Christensen J, Cotmore SF & Tattersall P (1997) A novel cellular site-specific DNA-binding protein cooperates with the viral NS1 polypeptide to initiate parvovirus DNA replication. *J Virol* 71(2):1405-1416.

Christensen J, Pedersen M, Aasted B & Alexandersen S (1995b) Purification and characterization of the major nonstructural protein (NS-1) of Aleutian mink disease parvovirus. *J Virol* 69(3):1802-1809.

Christensen J, Storgaard T, Viuff B, Aasted B & Alexandersen S (1993) Comparison of promoter activity in Aleutian mink disease parvovirus, minute virus of mice, and canine parvovirus: possible role of weak promoters in the pathogenesis of Aleutian mink disease parvovirus infection. *J Virol* 67(4):1877-1886.

Christensen J & Tattersall P (2002) Parvovirus initiator protein NS1 and RPA coordinate replication fork progression in a reconstituted DNA replication system. *J Virol* 76(13):6518-6531.

Chu DK, Poon LL, Chiu SS, Chan KH, Ng EM, Bauer I, Cheung TK, Ng IH, Guan Y, Wang D & Peiris JS (2012) Characterization of a novel gyrovirus in human stool and chicken meat. *J Clin Virol* 55(3):209-213.

Claessens JA, Schrier CC, Mockett AP, Jagt EH & Sondermeijer PJ (1991) Molecular cloning and sequence analysis of the genome of chicken anaemia agent. *J Gen Virol* 72 (Pt 8):2003-2006.

Clemens KE, Cerutis DR, Burger LR, Yang CQ & Pintel DJ (1990) Cloning of minute virus of mice cDNAs and preliminary analysis of individual viral proteins expressed in murine cells. *J Virol* 64(8):3967-3973.

Cortey M, Olvera A, Grau-Roma L & Segalés J (2011) Further comments on porcine circovirus type 2 (PCV2) genotype definition and nomenclature. *Veterinary Microbiology* 149(3–4):522-523.

Cossons N, Faust EA & Zannis-Hadjopoulos M (1996) DNA polymerase delta-dependent formation of a hairpin structure at the 5' terminal palindrome of the minute virus of mice genome. *Virology* 216(1):258-264.

Cotmore SF, Christensen J, Nuesch JP & Tattersall P (1995a) The NS1 polypeptide of the murine parvovirus minute virus of mice binds to DNA sequences containing the motif [ACCA]2-3. *J Virol* 69(3):1652-1660.

Cotmore SF, D'Abramo A M, Jr., Ticknor CM & Tattersall P (1999) Controlled conformational transitions in the MVM virion expose the VP1 N-terminus and viral genome without particle disassembly. *Virology* 254(1):169-181.

Cotmore SF, Gunther M & Tattersall P (1989a) Evidence for a ligation step in the DNA replication of the autonomous parvovirus minute virus of mice. *J Virol* 63(2):1002-1006.

Cotmore SF & Tattersall P (1987) The autonomously replicating parvoviruses of vertebrates. *Adv Virus Res* 33:91-174.

Cotmore SF & Tattersall P (1989b) A genome-linked copy of the NS-1 polypeptide is located on the outside of infectious parvovirus particles. *J Virol* 63(9):3902-3911.

Cotmore SF & Tattersall P (1990) Alternate splicing in a parvoviral nonstructural gene links a common amino-terminal sequence to downstream domains which confer radically different localization and turnover characteristics. *Virology* 177(2):477-487.

Cotmore SF & Tattersall P (1995b) DNA replication in the autonomous parvoviruses *Sem. Virol.* 6:271-281.

Cotmore SF & Tattersall P (1998) High-mobility group 1/2 proteins are essential for initiating rolling-circle-type DNA replication at a parvovirus hairpin origin. *J Virol* 72(11):8477-8484.

Cotmore SF & Tattersall P (2005) Genome packaging sense is controlled by the efficiency of the nick site in the right-end replication origin of parvoviruses minute virus of mice and LuIII. *J Virol* 79(4):2287-2300.

Cotmore SF & Tattersall P (2006) A rolling-hairpin strategy: basic mechanisms of DNA replication in the parvovirus. . *Parvoviruses*, (Ed.) IMLD (Édit.) Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.

Croizier L, Jousset FX, Veyrunes JC, Lopez-Ferber M, Bergoin M & Croizier G (2000) Protein requirements for assembly of virus-like particles of *Junonia coenia* densovirus in insect cells. *J Gen Virol* 81(Pt 6):1605-1613.

Danen-Van Oorschot AA, Fischer DF, Grimbergen JM, Klein B, Zhuang S, Falkenburg JH, Backendorf C, Quax PH, Van der Eb AJ & Noteborn MH (1997) Apoptin induces apoptosis in human transformed and malignant cells but not in normal cells. *Proc Natl Acad Sci U S A* 94(11):5843-5847.

Dayaram A, Potter KA, Moline AB, Rosenstein DD, Marinov M, Thomas JE, Breitbart M, Rosario K, Arguello-Astorga GR & Varsani A (2013) High global diversity of cycloviruses amongst dragonflies. *J Gen Virol* 94(Pt 8):1827-1840.

Dean FB, Nelson JR, Giesler TL & Lasken RS (2001) Rapid amplification of plasmid and phage DNA using Phi 29 DNA polymerase and multiply-primed rolling circle amplification. *Genome Res* 11(6):1095-1099.

Deleu L, Fuks F, Spitkovsky D, Horlein R, Faisst S & Rommelaere J (1998) Opposite transcriptional effects of cyclic AMP-responsive elements in confluent or p27KIP-overexpressing cells versus serum-starved or growing cells. *Mol Cell Biol* 18(1):409-419.

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(1-2):114-121.

Dhar AK, Kaizer KN, Betz YM, Harvey TN & Lakshman DK (2011) Identification of the core sequence elements in *Penaeus stylirostris* densovirus promoters. *Virus Genes* 43(3):367-375.

Dhar AK, Kaizer KN & Lakshman DK (2010) Transcriptional analysis of *Penaeus stylirostris* densovirus genes. *Virology* 402(1):112-120.

Dhar AK, Lakshman DK, Natarajan S, Allnutt FC & van Beek NA (2007) Functional characterization of putative promoter elements from infectious hypodermal and hematopoietic necrosis virus (IHHNV) in shrimp and in insect and fish cell lines. *Virus Res* 127(1):1-8.

Ding C, Urabe M, Bergoin M & Kotin RM (2002) Biochemical characterization of *Junonia coenia* densovirus nonstructural protein NS-1. *J Virol* 76(1):338-345.

Direkbusarakom S, Yoshimizu M, Ezura Y, Ruangpan L & Danayadol Y (1998) Vibrio spp., the dominant flora in shrimp hatchery against some fish pathogenic viruses. *J Mar Biotechnol* 6(4):266-267.

Doerig C, Hirt B, Antonietti JP & Beard P (1990) Nonstructural protein of parvoviruses B19 and minute virus of mice controls transcription. *J Virol* 64(1):387-396.

dos Santos HF, Knak MB, de Castro FL, Slongo J, Ritterbusch GA, Klein TA, Esteves PA, Silva AD, Trevisol IM, Claassen EA, Cornelissen LA, Lovato M, Franco AC, Roehe PM & Rijsewijk FA (2012) Variants of the recently discovered avian gyrovirus 2 are detected in Southern Brazil and The Netherlands. *Vet Microbiol* 155(2-4):230-236.

Douglas AJ, Phenix K, Mawhinney KA, Todd D, Mackie DP & Curran WL (1995) Identification of a 24 kDa protein expressed by chicken anaemia virus. *J Gen Virol* 76 (Pt 7):1557-1562.

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 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(4):1375-1380.

Eichwald V, Daeffler L, Klein M, Rommelaere J & Salome N (2002) The NS2 proteins of parvovirus minute virus of mice are required for efficient nuclear egress of progeny virions in mouse cells. *J Virol* 76(20):10307-10319.

El-Far M, Li Y, Fediere G, Abol-Ela S & Tijssen P (2004) Lack of infection of vertebrate cells by the densovirus from the maize worm *Mythimna loreyi* (MlDNV). *Virus Res* 99(1):17-24.

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(12):7024.

Ellis J, Hassard L, Clark E, Harding J, Allan G, Willson P, Strokappe J, Martin K, McNeilly F, Meehan B, Todd D & Haines D (1998) Isolation of circovirus from lesions of pigs with postweaning multisystemic wasting syndrome. *Can Vet J* 39(1):44-51.

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(47):17148-17153.

Fediere G (1983) *Recherche sur des viroses de lépidoptères Limacodidae ravageurs de palmacées*. Thèse de Doctorat (Univ. Sci. Techn., Languedoc, Monpéllier).

Fediere G, El-Far M, Li Y, Bergoin M & Tijssen P (2004) Expression strategy of densonucleosis virus from *Mythimna loreyi*. *Virology* 320(1):181-189.

Fediere G, Li Y, Zadori Z, Szelei J & Tijssen P (2002) Genome organization of *Casphalia 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. *Anal Biochem* 385(1):171-173.

Finsterbusch T, Steinfeldt T, Caliskan R & Mankertz A (2005) Analysis of the subcellular localization of the proteins Rep, Rep' and Cap of porcine circovirus type 1. *Virology* 343(1):36-46.

Flegel TW (1997) Major viral diseases of the black tiger prawn (*Penaeus monodon*) in Thailand *World J. Microbiol. Biotech.* 13:433-442.

Flegel TW (2006) Detection of major penaeid shrimp viruses in Asia, a historical perspective with emphasis on Thailand. *Aquaculture* 258(1–4):1-33.

Flegel TW (2012) Historic emergence, impact and current status of shrimp pathogens in Asia. *J Invertebr Pathol* 110(2):166-173.

Fraser CA & Owens L (1996) Spawner-isolated mortality virus from Australian *Penaeus monodon*. *Dis. Aquat. Org.* 27:141-148.

Gagnon CA, Music N, Fontaine G, Tremblay D & Harel J (2010) Emergence of a new type of porcine circovirus in swine (PCV): A type 1 and type 2 PCV recombinant. *Veterinary Microbiology* 144(1–2):18-23.

Gagnon CA, Tremblay D, Tijssen P, Venne MH, Houde A & Elahi SM (2007) The emergence of porcine circovirus 2b genotype (PCV-2b) in swine in Canada. *Canadian Veterinary Journal* 48(8):811-819.

Gangnonngiw W, Kiatpathomchai W, Sriurairatana S, Laisutisan K, Chuchird N, Limsuwan C & Flegel TW (2009) Parvo-like virus in the hepatopancreas of freshwater prawns Macrobrachium rosenbergii cultivated in Thailand. *Diseases of Aquatic Organisms* 85(3):167-173.

Ge X, Li J, Peng C, Wu L, Yang X, Wu Y, Zhang Y & Shi Z (2011) Genetic diversity of novel circular ssDNA viruses in bats in China. *J Gen Virol* 92(Pt 11):2646-2653.

Gibbs MJ & Weiller GF (1999) Evidence that a plant virus switched hosts to infect a vertebrate and then recombined with a vertebrate-infecting virus. *Proc Natl Acad Sci U S A* 96(14):8022-8027.

Gillespie J, Opriessnig T, Meng XJ, Pelzer K & Buechner-Maxwell V (2009) Porcine circovirus type 2 and porcine circovirus-associated disease. *J Vet Intern Med* 23(6):1151-1163.

Giraud C, Devauchelle G & Bergoin M (1992) The densovirus of *Junonia coenia* (Jc DNV) as an insect cell expression vector. *Virology* 186(1):207-218.

Gorbalenya AE & Koonin EV (1993) Helicases: amino acid sequence comparisons and structure-function relationships. *Current Opinion in Structural Biology* 3(3):419-429.

Gorbalenya AE, Koonin EV, Donchenko AP & Blinov VM (1989) Two related superfamilies of putative helicases involved in replication, recombination, repair and expression of DNA and RNA genomes. *Nucleic Acids Res* 17(12):4713-4730.

Goryo M, Sugimura H, Matsumoto S, Umemura T & Itakura C (1985) Isolation of an agent inducing chicken anaemia. *Avian Pathol* 14(4):483-496.

Greene EL & Karasaki S (1965) Physical and chemical properties of H-1 virus. II. Partial purification. *Proc Soc Exp Biol Med* 119(4):918-922.

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

Guo LJ, Lu YH, Wei YW, Huang LP & Liu CM (2010) Porcine circovirus type 2 (PCV2): genetic variation and newly emerging genotypes in China. *Virol J* 7:273.

Gutierrez C (1999) Geminivirus DNA replication. *Cell Mol Life Sci* 56(3-4):313-329.

Halami MY, Nieper H, Muller H & Johne R (2008) Detection of a novel circovirus in mute swans (Cygnus olor) by using nested broad-spectrum PCR. *Virus Res* 132(1-2):208-212.

Hamel AL, Lin LL & Nayar GP (1998) Nucleotide sequence of porcine circovirus associated with postweaning multisystemic wasting syndrome in pigs. *J Virol* 72(6):5262-5267.

Handa A, Dickstein B, Young NS & Brown KE (2000) Prevalence of the newly described human circovirus, TTV, in United States blood donors. *Transfusion* 40(2):245-251.

Harris RE, Coleman PH & Morahan PS (1974) Stability of minute virus of mice to chemical and physical agents. *Appl Microbiol* 28(3):351-354.

Hattermann K, Schmitt C, Soike D & Mankertz A (2003) Cloning and sequencing of Duck circovirus (DuCV). *Arch Virol* 148(12):2471-2480.

Hayakawa T, Kojima K, Nonaka K, Nakagaki M, Sahara K, Asano S, Iizuka T & Bando H (2000) Analysis of proteins encoded in the bipartite genome of a new type of parvo-like virus isolated from silkworm - structural protein with DNA polymerase motif. *Virus Res* 66(1):101-108.

Hazreen Nita MK, Kua BC, Bhassu S & Othman RY (2012) Detection and genetic profiling of infectious hypodermal and haematopoietic necrosis virus (IHHNV) infections in wild berried freshwater prawn, Macrobrachium rosenbergii collected for hatchery production. *Mol Biol Rep* 39(4):3785-3790.

Heath L, Williamson AL & Rybicki EP (2006) The capsid protein of beak and feather disease virus binds to the viral DNA and is responsible for transporting the replication-associated protein into the nucleus. *J Virol* 80(14):7219-7225.

Hsu JY, Juven-Gershon T, Marr MT, 2nd, Wright KJ, Tjian R & Kadonaga JT (2008) TBP, Mot1, and NC2 establish a regulatory circuit that controls DPE-dependent versus TATA-dependent transcription. *Genes Dev* 22(17):2353-2358.

Hu Y, Zheng J, Iizuka T & Bando H (1994) A densovirus newly isolated from the smoky-brown cockroach Periplaneta fuliginosa. *Arch Virol* 138(3-4):365-372.

Huynh OT, Pham HT, Yu Q & Tijssen P (2012) *Pseudoplusia includens* densovirus genome organization and expression strategy. *J Virol* 86(23):13127-13128.

Ilyina TV & Koonin EV (1992) Conserved sequence motifs in the initiator proteins for rolling circle DNA replication encoded by diverse replicons from eubacteria, eucaryotes and archaebacteria. *Nucleic Acids Res* 20(13):3279-3285.

Im DS & Muzyczka N (1990) The AAV origin binding protein Rep68 is an ATP-dependent site-specific endonuclease with DNA helicase activity. *Cell* 61(3):447-457.

Islam MR, Johne R, Raue R, Todd D & Muller H (2002) Sequence analysis of the full-length cloned DNA of a chicken anaemia virus (CAV) strain from Bangladesh: evidence for genetic grouping of CAV strains based on the deduced VP1 amino acid sequences. *J Vet Med B Infect Dis Vet Public Health* 49(7):332-337.

Iyer LM, Aravind L & Koonin EV (2001) Common origin of four diverse families of large eukaryotic DNA viruses. *J Virol* 75(23):11720-11734.

Iyer LM, Leipe DD, Koonin EV & Aravind L (2004) Evolutionary history and higher order classification of AAA+ ATPases. *J Struct Biol* 146(1-2):11-31.

James JA, Escalante CR, Yoon-Robarts M, Edwards TA, Linden RM & Aggarwal AK (2003) Crystal structure of the SF3 helicase from adeno-associated virus type 2. *Structure* 11(8):1025-1035.

Jensen MA, Fukushima M & Davis RW (2010) DMSO and betaine greatly improve amplification of GC-rich constructs in de novo synthesis. *PLoS One* 5(6):e11024.

Jeurissen SH, Wagenaar F, Pol JM, van der Eb AJ & Noteborn MH (1992) Chicken anemia virus causes apoptosis of thymocytes after in vivo infection and of cell lines after in vitro infection. *J Virol* 66(12):7383-7388.

Johne R, Muller H, Rector A, van Ranst M & Stevens H (2009) Rolling-circle amplification of viral DNA genomes using phi29 polymerase. *Trends Microbiol* 17(5):205-211.

Jourdan M, Jousset FX, 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.

Jousset FX, Baquerizo E & Bergoin M (2000) A new densovirus isolated from the mosquito *Culex pipiens* (Diptera: culicidae). *Virus Res* 67(1):11-16.

Jousset FX, Barreau C, Boublik Y & Cornet M (1993) A parvo-like virus persistently infecting a C6/36 clone of Aedes albopictus mosquito cell line and pathogenic for Aedes aegypti larvae. *Virus Res* 29(2):99-114.

Jung A, Ruckert S, Frank P, Brabletz T & Kirchner T (2002) 7-Deaza-2'-deoxyguanosine allows PCR and sequencing reactions from CpG islands. *Mol Pathol* 55(1):55-57.

Kamada K, Kuroishi A, Kamahora T, Kabat P, Yamaguchi S & Hino S (2006) Spliced mRNAs detected during the life cycle of Chicken anemia virus. *J Gen Virol* 87(Pt 8):2227-2233.

Kapelinskaya TV, Martynova EU, Schal C & Mukha DV (2011) Expression strategy of densonucleosis virus from the German cockroach, *Blattella germanica*. *J Virol* 85(22):11855-11870.

Kapoor A, Dubovi EJ, Henriquez-Rivera JA & Lipkin WI (2012) Complete genome sequence of the first canine circovirus. *J Virol* 86(12):7018.

Kaufmann B, Bowman VD, Li Y, Szelei J, Waddell PJ, Tijssen P & Rossmann MG (2010) Structure of *Penaeus stylirostris* densovirus, a shrimp pathogen. *J Virol* 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. *J Virol* 85(10):4691-4697.

Kaufmann B, Simpson AA & Rossmann MG (2004) The structure of human parvovirus B19. *Proc Natl Acad Sci U S A* 101(32):11628-11633.

Kautsky N, Rönnbäck P, Tedengren M & Troell M (2000) Ecosystem perspectives on management of disease in shrimp pond farming. *Aquaculture* 191(1–3):145-161.

Kimmick MW, Afanasiev BN, Beaty BJ & Carlson JO (1998) Gene expression and regulation from the p7 promoter of Aedes densonucleosis virus. *J Virol* 72(5):4364-4370.

Koch G, van Roozelaar DJ, Verschueren CA, van der Eb AJ & Noteborn MH (1995) Immunogenic and protective properties of chicken anaemia virus proteins expressed by baculovirus. *Vaccine* 13(8):763-770.

Kontou M, Govindasamy L, Nam HJ, Bryant N, Llamas-Saiz AL, Foces-Foces C, Hernando E, Rubio MP, McKenna R, Almendral JM & Agbandje-McKenna M (2005) Structural determinants of tissue tropism and in vivo pathogenicity for the parvovirus minute virus of mice. *J Virol* 79(17):10931-10943.

Koonin EV & Ilyina TV (1993) Computer-assisted dissection of rolling circle DNA replication. *Biosystems* 30(1-3):241-268.

Kouassi N, Peng J, Li Y, Cavallaro C, Veyrunes JC & Bergoin M (2007) Pathogenicity of Diatraea saccharalis Densovirus to host insect and characterization of its viral genome. *Virologica Sinica* 22(1):53-60.

Krupovic M, Ravantti JJ & Bamford DH (2009) Geminiviruses: a tale of a plasmid becoming a virus. *BMC Evol Biol* 9:112.

Kucharski TJ, Gamache I, Gjoerup O & Teodoro JG (2011) DNA damage response signaling triggers nuclear localization of the chicken anemia virus protein Apoptin. *J Virol* 85(23):12638-12649.

Kutach AK & Kadonaga JT (2000) The downstream promoter element DPE appears to be as widely used as the TATA box in Drosophila core promoters. *Mol Cell Biol* 20(13):4754-4764.

La Fauce KA, Elliman J & Owens L (2007) Molecular characterisation of hepatopancreatic parvovirus (PmergDNV) from Australian Penaeus merguiensis. *Virology* 362(2):397-403.

Laufs J, Traut W, Heyraud F, Matzeit V, Rogers SG, Schell J & Gronenborn B (1995) In vitro cleavage and joining at the viral origin of replication by the replication initiator protein of tomato yellow leaf curl virus. *Proc Natl Acad Sci U S A* 92(9):3879-3883.

Lefeuvre P, Lett JM, Varsani A & Martin DP (2009) Widely conserved recombination patterns among single-stranded DNA viruses. *J Virol* 83(6):2697-2707.

Li L, Kapoor A, Slikas B, Bamidele OS, Wang C, Shaukat S, Masroor MA, Wilson ML, Ndjango JB, Peeters M, Gross-Camp ND, Muller MN, Hahn BH, Wolfe ND, Triki H, Bartkus J, Zaidi SZ & Delwart E (2010) Multiple diverse circoviruses infect farm animals and are commonly found in human and chimpanzee feces. *J Virol* 84(4):1674-1682.

Li L, Shan T, Soji OB, Alam MM, Kunz TH, Zaidi SZ & Delwart E (2011) Possible cross-species transmission of circoviruses and cycloviruses among farm animals. *J Gen Virol* 92(Pt 4):768-772.

Li TW & Weeks KM (2006) Structure-independent and quantitative ligation of single-stranded DNA. *Anal Biochem* 349(2):242-246.

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(Pt 11):2821-2825.

Lightner DV (1996a) Epizootiology, distribution and the impact on international trade of two penaeid shrimp viruses in the Americas. *Rev Sci Tech* 15(2):579-601.

Lightner DV (1996b) *A handbook of Pathology and Diagnostic Procedures for Diseases of Penaeid Prawn.* 2 (Section 3) p

Lightner DV (1999) The penaeid shrimp viruses TSV, IHHNV, WSSV, and YHV: current status in the Americas, available diagnostic methods and management strategies. *J Appl Aquaculture* 9(2):27-52.

Lightner DV (2005) Biosecurity in Shrimp Farming: Pathogen Exclusion through Use of SPF Stock and Routine Surveillance. *Journal of the World Aquaculture Society* 36(3):229-248.

Lightner DV & Redman RM (1985) A parvo-like virus disease of penaeid shrimp. *Journal of Invertebrate Pathology* 45(1):47-53.

Lightner DV, Redman RM & Bell TA (1983) Infectious hypodermal and haematopoietic necrosis, a newly recognized virus disease of penaeid shrimp. . *J. Invertebr. Pathol 42* 42:62-70.

Lightner DV, Redman RM, Poulos BT, Nunan LM, Mari JL & Hasson KW (1997) Risk of spread of penaeid shrimp viruses in the Americas by the international movement of live and frozen shrimp. *Rev Sci Tech* 16(1):146-160.

Liu J, Chen I & Kwang J (2005) Characterization of a previously unidentified viral protein in porcine circovirus type 2-infected cells and its role in virus-induced apoptosis. *J Virol* 79(13):8262-8274.

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(19):10069-10078.

Liu Q, Tikoo SK & Babiuk LA (2001) Nuclear localization of the ORF2 protein encoded by porcine circovirus type 2. *Virology* 285(1):91-99.

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(Pt 1):93-102.

Lochelt M, Delius H & Kaaden OR (1989) A novel replicative form DNA of Aleutian disease virus: the covalently closed linear DNA of the parvoviruses. *J Gen Virol* 70 (Pt 5):1105-1116.

Lombardo E, Ramirez JC, Agbandje-McKenna M & Almendral JM (2000) A beta-stranded motif drives capsid protein oligomers of the parvovirus minute virus of mice into the nucleus for viral assembly. *J Virol* 74(8):3804-3814.

Lorincz M, Csagola A, Farkas SL, Szekely C & Tuboly T (2011) First detection and analysis of a fish circovirus. *J Gen Virol* 92(Pt 8):1817-1821.

Lorincz M, Dan A, Lang M, Csaba G, Toth AG, Szekely C, Csagola A & Tuboly T (2012) Novel circovirus in European catfish (Silurus glanis). *Arch Virol* 157(6):1173-1176.

Luo G, Xin S & Cai Z (2003) Role of the 5'-proximal stem-loop structure of the 5' untranslated region in replication and translation of hepatitis C virus RNA. *J Virol* 77(5):3312-3318.

Maggi F, Macera L, Focosi D, Vatteroni ML, Boggi U, Antonelli G, Eloit M & Pistello M (2012) Human gyrovirus DNA in human blood, Italy. *Emerg Infect Dis* 18(6):956-959.

Manivannan S, Otta SK & Karunasagar I (2002) Multiple viral infection in *Penaeus monodon* shrimp postlarvae in an Indian hatchery. *Dis Aquat Organ* 48(3):233-236.

Mankertz A, Caliskan R, Hattermann K, Hillenbrand B, Kurzendoerfer P, Mueller B, Schmitt C, Steinfeldt T & Finsterbusch T (2004) Molecular biology of Porcine circovirus: analyses of gene expression and viral replication. *Vet Microbiol* 98(2):81-88.

Mankertz A, Hattermann K, Ehlers B & Soike D (2000) Cloning and sequencing of columbid circovirus (coCV), a new circovirus from pigeons. *Arch Virol* 145(12):2469-2479.

Mankertz A & Hillenbrand B (2001) Replication of porcine circovirus type 1 requires two proteins encoded by the viral rep gene. *Virology* 279(2):429-438.

Markowski-Grimsrud CJ & Schat KA (2003) Infection with chicken anaemia virus impairs the generation of pathogen-specific cytotoxic T lymphocytes. *Immunology* 109(2):283-294.

Meehan BM, Creelan JL, McNulty MS & Todd D (1997) Sequence of porcine circovirus DNA: affinities with plant circoviruses. *J Gen Virol* 78 (Pt 1):221-227.

Meehan BM, McNeilly F, Todd D, Kennedy S, Jewhurst VA, Ellis JA, Hassard LE, Clark EG, Haines DM & Allan GM (1998) Characterization of novel circovirus DNAs associated with wasting syndromes in pigs. *J Gen Virol* 79 (Pt 9):2171-2179.

Meng G, Zhang X, Plevka P, Yu Q, Tijssen P & Rossmann MG (2013) The structure and host entry of an invertebrate parvovirus. *J Virol* 87(23):12523-12530.

Meynardier G, Matz G, Veyrunes JC & Bres N (1977) Virus de type densonucléose chez les orthoptères. *Ann. SOc. Entomol. Fr.* 13:487-493.

Meynardier G, Vago C, Plantevine G & Atger P (1964) Virose d'un type habituel chez le Lepidoptere, *Galleria mellonella. L. Rev. Zool. Agric. Appl.* 63:207-209.

Miller CL & Pintel DJ (2002) Interaction between parvovirus NS2 protein and nuclear export factor Crm1 is important for viral egress from the nucleus of murine cells. *J Virol* 76(7):3257-3266.

Mir MA & Panganiban AT (2010) The triplet repeats of the Sin Nombre hantavirus 5' untranslated region are sufficient in cis for nucleocapsid-mediated translation initiation. *J Virol* 84(17):8937-8944.

Miyata H, Tsunoda H, Kazi A, Yamada A, Khan MA, Murakami J, Kamahora T, Shiraki K & Hino S (1999) Identification of a novel GC-rich 113-nucleotide region to complete the circular, single-stranded DNA genome of TT virus, the first human circovirus. *J Virol* 73(5):3582-3586.

Mukha DV, Chumachenko AG, Dykstra MJ, Kurtti TJ & Schal C (2006) Characterization of a new densovirus infecting the German cockroach, *Blattella germanica*. *J Gen Virol* 87(Pt 6):1567-1575.

Multeau C, Froissart R, Perrin A, Castelli I, Casartelli M & Ogliastro M (2012) Four amino acids of an insect densovirus capsid determine midgut tropism and virulence. *J Virol* 86(10):5937-5941.

Musso M, Bocciardi R, Parodi S, Ravazzolo R & Ceccherini I (2006) Betaine, dimethyl sulfoxide, and 7-deaza-dGTP, a powerful mixture for amplification of GC-rich DNA sequences. *J Mol Diagn* 8(5):544-550.

Naccache SN, Hackett J, Jr., Delwart EL & Chiu CY (2014) Concerns over the origin of NIH-CQV, a novel virus discovered in Chinese patients with seronegative hepatitis. *Proc Natl Acad Sci U S A* 111(11):E976.

Naeger LK, Cater J & Pintel DJ (1990) The small nonstructural protein (NS2) of the parvovirus minute virus of mice is required for efficient DNA replication and infectious virus production in a cell-type-specific manner. *J Virol* 64(12):6166-6175.

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.

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(23):14758-14763.

Nawagitgul P, Morozov I, Bolin SR, Harms PA, Sorden SD & Paul PS (2000) Open reading frame 2 of porcine circovirus type 2 encodes a major capsid protein. *J Gen Virol* 81(Pt 9):2281-2287.

Ng TF, Willner DL, Lim YW, Schmieder R, Chau B, Nilsson C, Anthony S, Ruan Y, Rohwer F & Breitbart M (2011) Broad surveys of DNA viral diversity obtained through viral metagenomics of mosquitoes. *PLoS One* 6(6):e20579.

Niagro FD, Forsthoefel AN, Lawther RP, Kamalanathan L, Ritchie BW, Latimer KS & Lukert PD (1998) Beak and feather disease virus and porcine circovirus genomes: intermediates between the geminiviruses and plant circoviruses. *Arch Virol* 143(9):1723-1744.

Nogueira EO, A JPF, Martins Soares R, Luiz Durigon E, Lazzarin S & Brentano L (2007) Genome sequencing analysis of Brazilian chicken anemia virus isolates that lack MSB-1 cell culture tropism. *Comp Immunol Microbiol Infect Dis* 30(2):81-96.

Noteborn MH, de Boer GF, van Roozelaar DJ, Karreman C, Kranenburg O, Vos JG, Jeurissen SH, Hoeben RC, Zantema A, Koch G & et al. (1991) Characterization of cloned chicken anemia virus DNA that contains all elements for the infectious replication cycle. *J Virol* 65(6):3131-3139.

Noteborn MH & Koch G (1995) Chicken anaemia virus infection: molecular basis of pathogenicity. *Avian Pathol* 24(1):11-31.

Noteborn MH, Kranenburg O, Zantema A, Koch G, de Boer GF & van der Eb AJ (1992) Transcription of the chicken anemia virus (CAV) genome and synthesis of its 52-kDa protein. *Gene* 118(2):267-271.

Noteborn MH, Todd D, Verschueren CA, de Gauw HW, Curran WL, Veldkamp S, Douglas AJ, McNulty MS, van der EA & Koch G (1994) A single chicken anemia virus protein induces apoptosis. *J Virol* 68(1):346-351.

Noteborn MH, Verschueren CA, Koch G & Van der Eb AJ (1998) Simultaneous expression of recombinant baculovirus-encoded chicken anaemia virus (CAV) proteins VP1 and VP2 is required for formation of the CAV-specific neutralizing epitope. *J Gen Virol* 79 (Pt 12):3073-3077.

Nuesch JP, Cotmore SF & Tattersall P (1995) Sequence motifs in the replicator protein of parvovirus MVM essential for nicking and covalent attachment to the viral origin: identification of the linking tyrosine. *Virology* 209(1):122-135.

Nuesch JP & Tattersall P (1993) Nuclear targeting of the parvoviral replicator molecule NS1: evidence for self-association prior to nuclear transport. *Virology* 196(2):637-651.

O'Neill KC, Shen HG, Lin K, Hemann M, Beach NM, Meng XJ, Halbur PG & Opriessnig T (2011) Studies on porcine circovirus type 2 vaccination of 5-day-old piglets. *Clin Vaccine Immunol* 18(11):1865-1871.

Opriessnig T, Meng XJ & Halbur PG (2007) Porcine circovirus type 2 associated disease: update on current terminology, clinical manifestations, pathogenesis, diagnosis, and intervention strategies. *J Vet Diagn Invest* 19(6):591-615.

Orozco BM & Hanley-Bowdoin L (1998) Conserved sequence and structural motifs contribute to the DNA binding and cleavage activities of a geminivirus replication protein. *J Biol Chem* 273(38):24448-24456.

Oshima K, Kakizawa S, Nishigawa H, Kuboyama T, Miyata S, Ugaki M & Namba S (2001) A plasmid of phytoplasma encodes a unique replication protein having both plasmid- and virus-like domains: clue to viral ancestry or result of virus/plasmid recombination? *Virology* 285(2):270-277.

Owens L, Haqshenas G, McElnea C & Coelen R (1998) Putative spawner-isolated mortality virus associated with mid-crop mortality syndrome in farmed *Penaeus monodon* from northern Australia. *Dis Aquat Organ* 34(3):177-185.

Owens L & McElnea C (2000) Natural infection of the redclaw crayfish Cherax quadricarinatus with presumptive spawner-isolated mortality virus. *Dis Aquat Organ* 40(3):219-223.

Padilla-Rodriguez M, Rosario K & Breitbart M (2013) Novel cyclovirus discovered in the Florida woods cockroach Eurycotis floridana (Walker). *Arch Virol* 158(6):1389-1392.

Parrish CR, Holmes EC, Morens DM, Park EC, Burke DS, Calisher CH, Laughlin CA, Saif LJ & Daszak P (2008) Cross-species virus transmission and the emergence of new epidemic diseases. *Microbiol Mol Biol Rev* 72(3):457-470.

Parrish CR & Kawaoka Y (2005) The origins of new pandemic viruses: the acquisition of new host ranges by canine parvovirus and influenza A viruses. *Annu Rev Microbiol* 59:553-586.

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.

Peters MA, Crabb BS, Washington EA & Browning GF (2006) Site-directed mutagenesis of the VP2 gene of Chicken anemia virus affects virus replication, cytopathology and host-cell MHC class I expression. *J Gen Virol* 87(Pt 4):823-831.

Peters MA, Jackson DC, Crabb BS & Browning GF (2002) Chicken anemia virus VP2 is a novel dual specificity protein phosphatase. *J Biol Chem* 277(42):39566-39573.

Pham HT, Bergoin M & Tijssen P (2013a) *Acheta domesticus* Volvovirus, a Novel Single-Stranded Circular DNA Virus of the House Cricket. *Genome Announc* 1(2):e0007913.

Pham HT, Yu Q, Bergoin M & Tijssen P (2013b) A Novel Ambisense Densovirus, *Acheta domesticus* Mini Ambidensovirus, from Crickets. *Genome Announc* 1(6).

Phan TG, Kapusinszky B, Wang C, Rose RK, Lipton HL & Delwart EL (2011) The fecal viral flora of wild rodents. *PLoS Pathog* 7(9):e1002218.

Phenix KV, Meehan BM, Todd D & McNulty MS (1994) Transcriptional analysis and genome expression of chicken anaemia virus. *J Gen Virol* 75 (Pt 4):905-909.

Poole GM (1972) Stability of a modified, live panleucopenia virus stored in liquid phase. *Appl Microbiol* 24(4):663-664.

Pullen SS & Friesen PD (1995) The CAGT motif functions as an initiator element during early transcription of the baculovirus transregulator ie-1. *J Virol* 69(6):3575-3583.

Purnell BA, Emanuel PA & Gilmour DS (1994) TFIID sequence recognition of the initiator and sequences farther downstream in Drosophila class II genes. *Genes Dev* 8(7):830-842.

Pyrc K, Jebbink MF, Berkhout B & van der Hoek L (2008) Detection of new viruses by VIDISCA. Virus discovery based on cDNA-amplified fragment length polymorphism. *Methods Mol Biol* 454:73-89.

Qiu J, Cheng F & Pintel D (2007) The abundant R2 mRNA generated by aleutian mink disease parvovirus is tricistronic, encoding NS2, VP1, and VP2. *J Virol* 81(13):6993-7000.

Qiu J, Yoto J, Tullis GE & Pintel D (Édit) (2006) *Parvovirus RNA processing strategies.* Hodder Arnold, London, United Kingdom. 253-274 p.

Ren X, Hoiczyk E & Rasgon JL (2008) Viral paratransgenesis in the malaria vector Anopheles gambiae. *PLoS Pathog* 4(8):e1000135.

Rhode SL, 3rd & Richard SM (1987) Characterization of the trans-activation-responsive element of the parvovirus H-1 P38 promoter. *J Virol* 61(9):2807-2815.

Rijsewijk FA, Dos Santos HF, Teixeira TF, Cibulski SP, Varela AP, Dezen D, Franco AC & Roehe PM (2011) Discovery of a genome of a distant relative of chicken anemia virus reveals a new member of the genus Gyrovirus. *Arch Virol* 156(6):1097-1100.

Riolobos L, Reguera J, Mateu MG & Almendral JM (2006) Nuclear transport of trimeric assembly intermediates exerts a morphogenetic control on the icosahedral parvovirus capsid. *J Mol Biol* 357(3):1026-1038.

Ritchie BW, Niagro FD, Lukert PD, Steffens WL, 3rd & Latimer KS (1989) Characterization of a new virus from cockatoos with psittacine beak and feather disease. *Virology* 171(1):83-88.

Roberts EJ, Buck KW & Coutts RH (1988) Characterization of potato yellow mosaic virus as a geminivirus with a bipartite genome. *Intervirology* 29(3):162-169.

Rosario K, Dayaram A, Marinov M, Ware J, Kraberger S, Stainton D, Breitbart M & Varsani A (2012a) Diverse circular ssDNA viruses discovered in dragonflies (Odonata: Epiprocta). *J Gen Virol* 93(Pt 12):2668-2681.

Rosario K, Duffy S & Breitbart M (2012b) A field guide to eukaryotic circular single-stranded DNA viruses: insights gained from metagenomics. *Arch Virol* 157(10):1851-1871.

Rosario K, Marinov M, Stainton D, Kraberger S, Wiltshire EJ, Collings DA, Walters M, Martin DP, Breitbart M & Varsani A (2011) Dragonfly cyclovirus, a novel single-stranded DNA virus discovered in dragonflies (Odonata: Anisoptera). *J Gen Virol* 92(Pt 6):1302-1308.

Rosenberger JK & Cloud SS (1989) The isolation and characterization of chicken anemia agent (CAA) from broilers in the United States. *Avian Dis* 33(4):707-713.

Rossmann MG & Johnson JE (1989) Icosahedral RNA virus structure. *Annu Rev Biochem* 58:533-573.

Royer C, Bossin H, Romane C, Bergoin M & Couble P (2001) High amplification of a densovirus-derived vector in larval and adult tissues of Drosophila. *Insect Mol Biol* 10(3):275-280.

Rusaini, La Fauce KA, Elliman J, Bowater RO & Owens L (2013) Endogenous Brevidensovirus-like elements in Cherax quadricarinatus: Friend or foe? *Aquaculture* 396–399(0):136-145.

Ryabov EV, Keane G, Naish N, Evered C & Winstanley D (2009) Densovirus induces winged morphs in asexual clones of the rosy apple aphid, Dysaphis plantaginea. *Proc Natl Acad Sci U S A* 106(21):8465-8470.

Saunders K, Lucy A & Stanley J (1992) RNA-primed complementary-sense DNA synthesis of the geminivirus African cassava mosaic virus. *Nucleic Acids Res* 20(23):6311-6315.

Sauvage V, Cheval J, Foulongne V, Gouilh MA, Pariente K, Manuguerra JC, Richardson J, Dereure O, Lecuit M, Burguiere A, Caro V & Eloit M (2011) Identification of the first human gyrovirus, a virus related to chicken anemia virus. *J Virol* 85(15):7948-7950.

Schat KA (2009) Chicken anemia virus. *Curr Top Microbiol Immunol* 331:151-183.

Schat KA & Skinner MA (2008) Avian immunosuppressive diseases and immune evasion. *Avian immunology*:299.

Schnepp BC, Jensen RL, Clark KR & Johnson PR (2009) Infectious molecular clones of adeno-associated virus isolated directly from human tissues. *J Virol* 83(3):1456-1464.

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.

Segales J (2012) Porcine circovirus type 2 (PCV2) infections: clinical signs, pathology and laboratory diagnosis. *Virus Res* 164(1-2):10-19.

Shackelton LA, Hoelzer K, Parrish CR & Holmes EC (2007) Comparative analysis reveals frequent recombination in the parvoviruses. *J Gen Virol* 88(Pt 12):3294-3301.

Shackelton LA, Parrish CR, Truyen U & Holmes EC (2005) High rate of viral evolution associated with the emergence of carnivore parvovirus. *Proc Natl Acad Sci U S A* 102(2):379-384.

Shan T, Li L, Simmonds P, Wang C, Moeser A & Delwart E (2011) The fecal virome of pigs on a high-density farm. *J Virol* 85(22):11697-11708.

Shike H, Dhar AK, Burns JC, Shimizu C, Jousset FX, Klimpel KR & Bergoin M (2000) Infectious Hypodermal and Hematopoietic Necrosis Virus of Shrimp Is Related to Mosquito Brevidensoviruses. *Virology* 277(1):167-177.

Shirk PD, Bossin H, Furlong RB & Gillett JL (2007) Regulation of *Junonia coenia* densovirus P9 promoter expression. *Insect Mol Biol* 16(5):623-633.

Simpson AA, Chipman PR, Baker TS, Tijssen P & Rossmann MG (1998) The structure of an insect parvovirus (*Galleria mellonella* densovirus) at 3.7 A resolution. *Structure* 6(11):1355-1367.

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(5):1189-1198.

Sivakumar VK, Sarathi M, Venkatesan C, Sivaraj A & Hameed ASS (2009) Experimental exposure of Artemia to Hepatopancreatic parvo-like Virus and Subsequent transmission to post-larvae of *Penaeus monodon*. *Journal of Invertebrate Pathology* 102(3):191-195.

Smale ST & Kadonaga JT (2003) The RNA polymerase II core promoter. *Annu Rev Biochem* 72:449-479.

Smits SL, Zijlstra EE, van Hellemond JJ, Schapendonk CM, Bodewes R, Schurch AC, Haagmans BL & Osterhaus AD (2013) Novel cyclovirus in human cerebrospinal fluid, Malawi, 2010-2011. *Emerg Infect Dis* 19(9).

Stanley J, Markham PG, Callis RJ & Pinner MS (1986) The nucleotide sequence of an infectious clone of the geminivirus beet curly top virus. *EMBO J* 5(8):1761-1767.

Steinfeldt T, Finsterbusch T & Mankertz A (2001) Rep and Rep' protein of porcine circovirus type 1 bind to the origin of replication in vitro. *Virology* 291(1):152-160.

Steinfeldt T, Finsterbusch T & Mankertz A (2006) Demonstration of nicking/joining activity at the origin of DNA replication associated with the rep and rep' proteins of porcine circovirus type 1. *J Virol* 80(13):6225-6234.

Steinfeldt T, Finsterbusch T & Mankertz A (2007) Functional analysis of cis- and trans-acting replication factors of porcine circovirus type 1. *J Virol* 81(11):5696-5704.

Styer EL & Hamm JJ (1991) Report of a densovirus in a commercial cricket operation in the southeastern United States. *Journal of invertebrate pathology* 58:283-285.

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.

Sukhumsirichart W, Wongteerasupaya C, Boonsaeng V, Panyim S, Sriurairatana S, B. W & Flegel TW (1999) Characterization and PCR detection of hepatopancreatic parvovirus (HPV) from *Penaeus monodon* in Thailand. *Diseases of Aquatic Organisms* 38(1):1-10.

Suto C, Kawamoto F & Kumada N (1979) A new virus isolated from the cockroach, *Periplaneta fuliginosa* (Serville). *Microbiol Immunol* 23(3):207-211.

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. *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.

Tan le V, van Doorn HR, Nghia HD, Chau TT, Tu le TP, de Vries M, Canuti M, Deijs M, Jebbink MF, Baker S, Bryant JE, Tham NT, NT BK, Boni MF, Loi TQ, Phuong le T, Verhoeven

JT, Crusat M, Jeeninga RE, Schultsz C, Chau NV, Hien TT, van der Hoek L, Farrar J & de Jong MD (2013) Identification of a new cyclovirus in cerebrospinal fluid of patients with acute central nervous system infections. *MBio* 4(3):e00231-00213.

Tan R & Frankel AD (1995) Structural variety of arginine-rich RNA-binding peptides. *Proc Natl Acad Sci U S A* 92(12):5282-5286.

Tang KF & Lightner DV (2006) Infectious hypodermal and hematopoietic necrosis virus (IHHNV)-related sequences in the genome of the black tiger prawn *Penaeus monodon* from Africa and Australia. *Virus Res* 118(1-2):185-191.

Tang KF, Poulos BT, Wang J, Redman RM, Shih HH & Lightner DV (2003) Geographic variations among infectious hypodermal and hematopoietic necrosis virus (IHHNV) isolates and characteristics of their infection. *Dis Aquat Organ.* 53(2):91-99.

Tattersall P (1972) Replication of the parvovirus MVM. I. Dependence of virus multiplication and plaque formation on cell growth. *J Virol* 10(4):586-590.

Teodoro JG, Heilman DW, Parker AE & Green MR (2004) The viral protein Apoptin associates with the anaphase-promoting complex to induce G2/M arrest and apoptosis in the absence of p53. *Genes Dev* 18(16):1952-1957.

Tijssen P, Agbandje-McKenna M, Almendral JM, Bergoin M, Flegel TW, Hedman K, Kleinschmidt JA, Li Y, Pintel DJ & Tattersall P (2011) Virus taxonomy: classification and nomenclature of viruses: Ninth report of the International Committee on Taxonomy of Viruses. . *Parvoviridae*, In King Amq AM, Carstens E and Lefkowitz Ej (Édit.) Elsevier, San Diego.

Tijssen P, Bando H, Li Y, Jousset FX, Zadori Z, Fediere G & El-Far M (2006a) Evolution of densoviruses. *Parvoviruses*, (Ed.) IMLD (Édit.) Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. p 55-66.

Tijssen P & Bergoin M (1995a) Densonucleosis viruses constitue an increasingly diversified among parvoviruses. *Sem. Virol.* 6:347-355.

Tijssen P & Bergoin* M (1995b) Densonucleosis viruses constitute an increasingly diversified subfamily among the parvoviruses. *Seminars in Virology* 6(5):347-355.

Tijssen P & Kurstak E (1981) Biochemical, biophysical, and biological properties of densonucleosis virus (parvovirus). III. common sequences of structural proteins. *J Virol* 37(1):17-23.

Tijssen P, Li Y, El-Far M, Szelei J, Letarte M & Zadori Z (2003) Organization and expression strategy of the ambisense genome of densonucleosis virus of *Galleria mellonella*. *J Virol* 77(19):10357-10365.

Tijssen P, Szelei J & Zadori Z (2006b) Phospholipase A2 domains in structural proteins of parvovirus. *Parvovirus*, (Ed.) IMLD (Édit.) Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. p 96-103.

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. *J Virol* 21(1):225-231.

Tijssen P, van den Hurk J & Kurstak E (1976) Biochemical, biophysical, and biological properties of densonucleosis virus. I. Structural proteins. *J Virol* 17(3):686-691.

Timchenko T, Katul L, Aronson M, Vega-Arreguin JC, Ramirez BC, Vetten HJ & Gronenborn B (2006) Infectivity of nanovirus DNAs: induction of disease by cloned genome components of Faba bean necrotic yellows virus. *J Gen Virol* 87(Pt 6):1735-1743.

Tischer I, Mields W, Wolff D, Vagt M & Griem W (1986) Studies on epidemiology and pathogenicity of porcine circovirus. *Arch Virol* 91(3-4):271-276.

Tischer I, Rasch R & Tochtermann G (1974) Characterization of papovavirus-and picornavirus-like particles in permanent pig kidney cell lines. *Zentralbl Bakteriol Orig A* 226(2):153-167.

Todd D, McNulty MS, Adair BM & Allan GM (2001) Animal circoviruses. *Adv Virus Res* 57:1-70.

Toolan HW, Dalldore G, Barclay M, Chandra S & Moore AE (1960) An Unidentified, Filtrable Agent Isolated from Transplanted Human Tumors. *Proc Natl Acad Sci U S A* 46(9):1256-1258.

Toro H, Ewald S & Hoerr FJ (2006) Serological evidence of chicken infectious anemia virus in the United States at least since 1959. *Avian Dis* 50(1):124-126.

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

Tullis G, Schoborg RV & Pintel DJ (1994) Characterization of the temporal accumulation of minute virus of mice replicative intermediates. *J Gen Virol* 75 (Pt 7):1633-1646.

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. *J Virol* 67(1):131-141.

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

Vega-Arreguin JC, Gronenborn B & Ramirez BC (2007) Arabidopsis thaliana is a host of the legume nanovirus Faba bean necrotic yellows virus. *Virus Res* 128(1-2):81-87.

Vega-Rocha S, Byeon IJ, Gronenborn B, Gronenborn AM & Campos-Olivas R (2007) Solution structure, divalent metal and DNA binding of the endonuclease domain from the replication initiation protein from porcine circovirus 2. *J Mol Biol* 367(2):473-487.

Verma B, Bhattacharyya S & Das S (2010) Polypyrimidine tract-binding protein interacts with coxsackievirus B3 RNA and influences its translation. *J Gen Virol* 91(Pt 5):1245-1255.

Vihinen-Ranta M, Wang D, Weichert WS & Parrish CR (2002) The VP1 N-terminal sequence of canine parvovirus affects nuclear transport of capsids and efficient cell infection. *J Virol* 76(4):1884-1891.

Walker PJ & Winton JR (2010) Emerging viral diseases of fish and shrimp. *Vet Res* 41(6):51.

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(Pt 8):2169-2173.

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

Ward TW, Jenkins MS, Afanasiev BN, Edwards M, Duda BA, Suchman E, Jacobs-Lorena M, Beaty BJ & Carlson JO (2001a) Aedes aegypti transducing densovirus pathogenesis and expression in Aedes aegypti and Anopheles gambiae larvae. *Insect Mol Biol* 10(5):397-405.

Ward TW, Kimmick MW, Afanasiev BN & Carlson JO (2001b) Characterization of the structural gene promoter of Aedes aegypti densovirus. *J Virol* 75(3):1325-1331.

Weis L & Reinberg D (1992) Transcription by RNA polymerase II: initiator-directed formation of transcription-competent complexes. *FASEB J* 6(14):3300-3309.

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.

Willwand K, Baldauf AQ, Deleu L, Mumtsidu E, Costello E, Beard P & Rommelaere J (1997) The minute virus of mice (MVM) nonstructural protein NS1 induces nicking of MVM DNA at a unique site of the right-end telomere in both hairpin and duplex conformations in vitro. *J Gen Virol* 78 (Pt 10):2647-2655.

Wilson GM, Jindal HK, Yeung DE, Chen W & Astell CR (1991) Expression of minute virus of mice major nonstructural protein in insect cells: purification and identification of ATPase and helicase activities. *Virology* 185(1):90-98.

Wolter S, Richards R & Armentrout RW (1980) Cell cycle-dependent replication of the DNA of minute virus of mice, a parvovirus. *Biochim Biophys Acta* 607(3):420-431.

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(16):10405-10410.

Xie Q & Chapman MS (1996) Canine parvovirus capsid structure, analyzed at 2.9 A resolution. *J Mol Biol* 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. *J Virol* 86(19):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. *Arch Virol* 144(11):2111-2124.

Yang B, Dong X, Cai D, Wang X, Liu Z, Hu Z, Wang H, Cao X, Zhang J & Hu Y (2008) Characterization of the promoter elements and transcription profile of *Periplaneta fuliginosa* densovirus nonstructural genes. *Virus Res* 133(2):149-156.

Yang B, Zhang J, Cai D, Li D, Chen W, Jiang H & Hu Y (2006) Biochemical characterization of *Periplaneta fuliginosa* densovirus non-structural protein NS1. *Biochem Biophys Res Commun* 342(4):1188-1196.

Yoon HS, Price DC, Stepanauskas R, Rajah VD, Sieracki ME, Wilson WH, Yang EC, Duffy S & Bhattacharya D (2011) Single-cell genomics reveals organismal interactions in uncultivated marine protists. *Science* 332(6030):714-717.

Young PJ, Jensen KT, Burger LR, Pintel DJ & Lorson CL (2002a) Minute virus of mice NS1 interacts with the SMN protein, and they colocalize in novel nuclear bodies induced by parvovirus infection. *J Virol* 76(8):3892-3904.

Young PJ, Jensen KT, Burger LR, Pintel DJ & Lorson CL (2002b) Minute virus of mice small nonstructural protein NS2 interacts and colocalizes with the Smn protein. *J Virol* 76(12):6364-6369.

Yozwiak NL, Skewes-Cox P, Stenglein MD, Balmaseda A, Harris E & DeRisi JL (2012) Virus identification in unknown tropical febrile illness cases using deep sequencing. *PLoS Negl Trop Dis* 6(2):e1485.

Yu Q, Fediere G, Abd-Alla A, Bergoin M & Tijssen P (2012a) Iteravirus-like genome organization of a densovirus from Sibine fusca Stoll. *J Virol* 86(16):8897-8898.

Yu Q, Hajek AE, Bergoin M & Tijssen P (2012b) Papilio polyxenes densovirus has an iteravirus-like genome organization. *J Virol* 86(17):9534-9535.

Yuasa N, Taniguchi T & Yoshida I (1979) Isolation and some characteristics of an agent inducing anemia in chicks. *Avian Dis* 23:366-385.

Zadori Z, Stefancsik R, Rauch T & Kisary J (1995) Analysis of the complete nucleotide sequences of goose and muscovy duck parvoviruses indicates common ancestral origin with adeno-associated virus 2. *Virology* 212(2):562-573.

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(2):291-302.

Zhai SL, Chen SN, Wei ZZ, Zhang JW, Huang L, Lin T, Yue C, Ran DL, Yuan SS, Wei WK & Long JX (2011) Co-existence of multiple strains of porcine circovirus type 2 in the same pig from China. *Virol J* 8:517.

Zhai YG, Lv XJ, Sun XH, Fu SH, Gong ZD, Fen Y, Tong SX, Wang ZX, Tang Q, Attoui H & Liang GD (2008) Isolation and characterization of the full coding sequence of a novel densovirus from the mosquito *Culex pipiens* pallens. *J Gen Virol* 89(Pt 1):195-199.

Zhang X, Xie Q, Ji J, Chang S, Liu J, Chen F, Ma J & Bee Y (2012) Complete genome sequence analysis of a recent chicken anemia virus isolate and comparison with a chicken anemia virus isolate from human fecal samples in China. *J Virol* 86(19):10896-10897.

Zhi N, Hu G, Wong S, Zhao K, Mao Q & Young NS (2014) Reply to Naccache et al: Viral sequences of NIH-CQV virus, a contamination of DNA extraction method. *Proc Natl Acad Sci U S A* 111(11):E977.

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

Zhuang SM, Shvarts A, van Ormondt H, Jochemsen AG, van der Eb AJ & Noteborn MH (1995) Apoptin, a protein derived from chicken anemia virus, induces p53-independent apoptosis in human osteosarcoma cells. *Cancer Res* 55(3):486-489.

Annexes

Annexe 1. Publication: Expression strategy of *Aedes albopictus* densovirus

JVI
Journals.ASM.org

Expression Strategy of *Aedes albopictus* Densovirus

Hanh T. Pham,[a] Françoise-Xavière Jousset,[b] Jonathan Perreault,[a] Hiroko Shike,[c] Jozsef Szelei,[a] Max Bergoin,[a] Peter Tijssen[a]

INRS-Institut Armand-Frappier, Laval, Québec, Canada[a]; Laboratoire de Pathologie Comparée, Université Montpellier II, Montpellier, France[b]; Division of Clinical Pathology, Milton S. Hershey Medical Center, Pennsylvania State University, Hershey, Pennsylvania, USA[c]

The transcription map of the *Aedes albopictus* densovirus (AalDNV) brevidensovirus was identified by Northern blotting, rapid amplification of cDNA ends (RACE) analysis, and RNase protection assays. AalDNV produced mRNAs of 3,359 (NS1), 3,345 (NS2), and 1,246 (VP) nucleotides. The two overlapping P7/7.4 NS promoters employed closely located alternate transcription initiation sites, positioned at either side of the NS1 initiation codon. All NS mRNAs coterminated with VP mRNA. All promoters, explored using luciferase assays, were functional in insect and human cell lines.

Invertebrate parvoviruses (densoviruses [DNVs]) are subdivided into four genera: *Densovirus*, *Pefudensovirus*, *Iteravirus*, and *Brevidensovirus* (1–3). Brevidensoviruses have a 4.1-kb single-stranded DNA (ssDNA) genome with three open reading frames (ORFs) on the same strand (about 790 [NS1], 360 [NS2], and 350 [VP] amino acids). The genome of *Aedes albopictus* densovirus (AalDNV) (GenBank accession no. NC_004285) has terminal, T-shaped hairpins (4). Brevidensoviruses were isolated from medically important mosquito vectors, such as *Aedes albopictus* (AalDNV) (4), *Aedes aegypti* (AaeDNV) (5), and *Anopheles gambiae* (AgDNV) (6). AalDNV was isolated from *Aedes albopictus* C6/36 cells (7) but is infectious for *Aedes aegypti* larvae (8, 9).

AaeDNV expression has been studied to some extent (10, 11). ORF_{NS1} of AaeDNV was reported to have a 57-amino-acid N-terminal extension compared to that of AalDNV (4, 5). β-Galactosidase fusion proteins with the three ORFs were enzymatically active, except for NS1 (12) unless NS1-β-gal was constructed downstream of the corresponding AalDNV AUG_{NS1} (11). Primer extension demonstrated that AaeDNV VP transcription started at nucleotide (nt) 2402 (10). Here, AalDNV transcription was analyzed by Northern hybridization, 5′ and 3′ rapid amplification of cDNA ends (RACE), amplicon mapping, and RNase protection assays (RPAs). AalDNV promoter activities were also determined.

AalDNV transcripts were obtained after transfection of pCR2.1-AalDNV (containing the AalDNV genome of 4,176 nt [4] between its EcoRI sites), using Lipofectamine (Invitrogen protocol), in permissive C6/36 cells (7) grown in RPMI medium supplemented with 5% fetal bovine serum (FBS). After 48 h, RNA was isolated using the Stratagene Absolutely RNA miniprep kit.

Northern blotting. ^{32}P-labeled RNA probes for Northern blots targeted overlapping domains of the NS1 and NS2 ORFs or VP ORF. PCR-amplified products (primers in Table 1) were transcribed *in vitro* with [α-^{32}P]UTP and T7 RNA polymerase (NEB) (13). Northern blots using 10 μg of total RNA (13, 14) revealed transcript sizes of 3.4 and 1.3 kb with the VP-specific probe and a 3.4-kb transcript with the NS-specific probe (Fig. 1A).

RACE experiments. The FirstChoice RLM RACE kit (Invitrogen) was used to identify the 5′ starts and 3′ ends of the polyadenylated transcripts. The 5′ adapter primer (IP) and the 3′-anchored primer (AP) (in the kit) were used in PCR with AalDNV-specific primers (Table 1), as shown in Fig. 1B. Amplicons obtained with A626R/IP (~325 bp), A497R/IP (~200 bp), A2740R/IP (~350 bp), A2540R/IP (~150 bp), A3162F/AP (~600 bp), and A3461F (~300 bp) (data not shown) were sequenced. Sequencing revealed that NS1 transcription started at nt 329 (AGTA), 6 nt upstream of AUG_{NS1}, and that VP transcription started at nt 2441 (CAGTCG), 158 nt upstream of AUG_{VP} (Fig. 1B) (sequence data not shown). Sequencing of the 300- and 600-bp amplicons showed a transcription termination position at 3680, 18 nt downstream of the polyadenylation signal at position 3662 (Fig. 1C) (sequence data not shown).

As is common for other densoviruses, NS2 could be translated from nt 411 by leaky scanning (13) on the NS1 transcript. The short untranscribed region of NS1 (UTR_{NS1}), in spite of a consensus AnnAUGG sequence for NS1 initiation, could favor it (13). A canonical initiator Inr_{NS2} sequence (CAGT) is located at nt 342. After TfiI digestion (NEB) at nt 338 of the A831R/IP PCR product, followed by a nested PCR (A626R/IP or A497R/IP) for specific amplification of putative NS2 transcripts, bands of 330 and 200 nt were produced (data not shown). Their sequencing revealed that NS2 transcription started at nt 343 (sequence data not shown). Thus, NS1 and NS2 AalDNV transcription starts were separated by 14 nt on either side of AUG_{NS1}.

Amplicon mapping. In order to confirm that NS and VP transcripts all coterminated at position 3680, primer A2380F (upstream of the VP transcript) and two reverse primers, A3650R (upstream) and A3744R (downstream), of the VP transcript end were designed for reverse transcription (RT)-PCR amplification (Table 1). Only a band of about 1,300 nt was obtained using the A2380F/A3650R set of primers for NS transcripts (Fig. 1C).

RNase protection assays. RPAs were employed to confirm NS1, NS2, and VP transcription starts. RPA probes for NS and VP transcripts were prepared by PCR (Table 1; Fig. 2A), and *in vitro* transcription as described for Northern blot probes. For positive controls, a sense RNA was generated by *in vitro* transcription and RPA was performed in parallel: for NS, two positive-control RNAs spanning from nt 329 to 439 (predicted for NS1) and nt 343 to 439 (predicted for NS2) were used, and for VP one from nt 2441 to 2542 was used. The probes and control RNA were purified from polyacrylamide gel and used in the RPA III kit (Invitrogen). Size markers were generated according to the 5′ RACE results. The NS

Received 9 May 2013 Accepted 18 June 2013
Published ahead of print 26 June 2013
Address correspondence to Peter Tijssen, peter.tijssen@iaf.inrs.ca.

doi:10.1128/JVI.01259-13

TABLE 1 Primers used for Northern blot probes, RACE, amplicon mapping, and RPAs

Primer name[a]	Sequence	Purpose
A639F	GCTCCAGAGCCTCTGAACAGCTTG	NS probe, Northern blot
A1345R (+T7 sequence)	TAATACGACTCACTATAGGGGTTCTGACTCTTGTGCTGTTTC	NS probe, Northern blot
A3106F	CTAGAAACAGTTGCAGCAACCGGAC	VP probe, Northern blot
A3509R (+T7 sequence)	TAATACGACTCACTATAGGCGTACTTGATATCTGAATTTCATG	VP probe, Northern blot
A3372F	AACTACAACATATGCCACGTCAG	3′ RACE
A3461F	ACAAGTTCCAGACGAAACAGG	3′ RACE
A497R	GTTCGTAATTGTTGGCATTCCT	5′ RACE of NS
A626R	GTGGGTAGATGTTATCAACGG	5′ RACE of NS
A831R	CTTGCCTGTGACCCGTTATTATCC	5′ RACE of NS
A2540R	GTGCGTTGTCTTCTTCTTCTATC	5′ RACE of VP
A2740R	GACCAAACATTACGGAAATGG	5′ RACE of VP
A3126F	CGGACCATTAGCACAACAAAC	3′ RACE
A2380F	GAGTATACAACACAGAGAAG	Amplicon mapping
A3650R	TCATAA GGCATACATGCTAC	Amplicon mapping
A3744R	TCTGTCGTGGACATTATCAG	Amplicon mapping
A272F (+UTS)	GCGATGAATGAACACTGAATCCACCACCACATGATCC	RPA NS probe
A329F (+T7 sequence)	TAATACGACTCACTATAGGAGTAGTATGGAATCAG	RPA-positive control for NS1
A343F (+T7 sequence)	TAATACGACTCACTATAGGGTCTGCAGTGAACATTCG	RPA-positive control for NS2
A439R	TCTCCTCCTGGATTTACACTG	RPA-positive control for NS1 and NS2
A439R (+T7 sequence)	TAATACGACTCACTATAGGTCTCCTCCTGGATTTACACTG	RPA NS probe
A2323F (+UTS)	GCGATGAATGAACACTGGCATATGAACGAAACCTCAC	RPA VP probe
A2441F (+T7 sequence)	TAATACGACTCACTATAGGAGTCGGCCACCAGGTCTTGTAG	RPA-positive control for VP
A2542R	ATGTGCGTTGTCTTCTTCTTC	RPA-positive control for VP
A2542R (+T7 sequence)	TAATACGACTCACTATAGGATGTGCGTTGTCTTCTTCTTC	RPA VP probe
A148F	TCCAATTGGAACACACGGAC	P7/7.4 promoter for AalDNV
A333R	CTACTGACTCTCCCTTC	P7/7.4 promoter for AalDNV
A2431F	CAAACTCATCAGTCGGCCAC	P60 promoter for AalDNV
A2597R	CCTCTGCTTCTTCTTTTGC	P60 promoter for AalDNV

[a] In the primer names, "UTS" represents the unspecific target sequence and "R" and "F" represent sense and antisense, respectively. The numbers in the primer names indicate the 5′ end of the primer sequence in AalDNV. The underlined sequences represent the UTS or T7 sequences.

probe protected the expected 96 and 110 nt of the NS transcript 5′ ends (Fig. 2B), confirming NS1 and NS2 mRNA start positions. Similarly, the VP probe confirmed the VP transcript 5′ extremity by protecting 101 nt (Fig. 2B). A protected band at 219 nt, slightly smaller than the entire probe (236 nt, including 17 nt of unspecific sequence), showed the coexistence of NS transcripts.

Promoter activity. NS (P7/7.4) and VP (P60) promoter regions were amplified by PCR (primers in Table 1) and cloned into

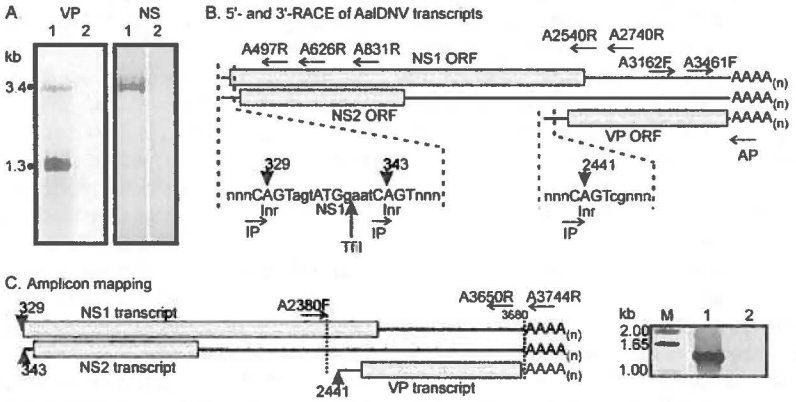


FIG 1 (A) Northern blot analysis of AalDNV transcription 48 h after transfection with vectors with virus-specific inserts (lanes 1) or without inserts (lanes 2). The VP probe also detected bands corresponding to the size of NS transcript indicating that VP and NS transcripts may coterminate. (B) Strategy of 5′ and 3′ RACE. Above the ORFs are the specific primers that were designed, and below are the ORFs and IP and AP primers from the Invitrogen kit (indicated with arrows). We took advantage of the TfiI restriction site between the putative Inr of NS1 and NS2 to distinguish between these transcripts. (After digestion, only amplicons from the second Inr would be obtained.) (C) Confirmation of termini of NS transcripts of AalDNV. The diagrams represent transcript map results by 5′ and 3′ RACE. Forward primers that do not recognize VP transcripts and reverse primers, both upstream and downstream of the VP transcript end (3′ RACE), were used in a PCR (indicated in the diagrams). As expected, only the inboard reverse primer (A3650R in lane 1) gave a product. M, markers; lane 2, A2380F/A3744R.

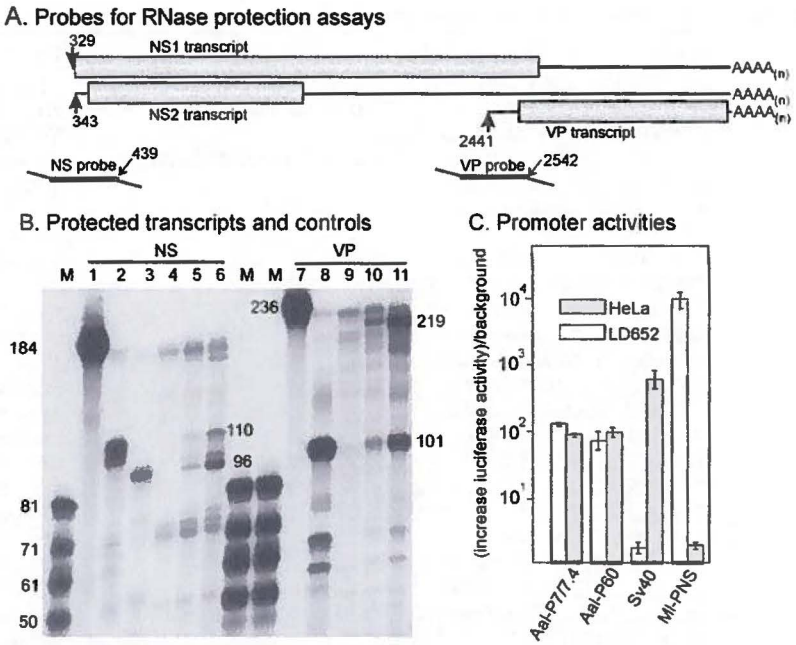


FIG 2 (A) Diagram of AalDNV expression and location of probes. (B) RNase protection assay of starts of NS and VP transcripts of AalDNV. Lanes: 1, NS probe with a specific length of 184 nt and short nonspecific extra terminal sequences; 2, positive control for NS1 with specific size obtained with RACE experiments; 3, positive control for NS2 with specific size obtained with RACE experiments; 4, total RNA from C6/36; 5, total RNA from AalDNV-infected C6/36 (15 μg); 6, total RNA from AalDNV-infected C6/36 (40 μg); 7, VP probe; 8, positive control for VP; 9, total RNA from C6/36; 10, total RNA from AalDNV-infected C6/36 (15 μg); 11, total RNA from AalDNV-infected C6/36 (40 μg). The band at 101 nt confirmed the VP start, and the band at 219 nt represented protection of the VP probe on the NS transcript. (C) Luciferase activity 40 h after transfection with different promoter constructs as fold increase over background (transfected pGL4 without insert). Inverse orientation of the promoters did not increase activity significantly over background. SV40 (not shown) and the insect virus MlDNV NS promoter (Ml-NSP) had a strong preference for cells from vertebrates and invertebrates, respectively. Surprisingly, AalDNV had significant activity in both types of cells.

EcoRV in pGL4.20, upstream of the luciferase gene (Promega; GenBank accession no. DQ188840) to estimate their functionality. Those with the reverse promoter orientation, shown by sequencing, served as controls. The *Mythimna loreyi* (MlDNV) NS promoter (an insect virus replicating in LD652 cells) (15) and the simian virus 40 (SV40) promoter were also cloned into pGL4.20 vectors to serve as positive controls. Transfection of LD652 using DOTAP (Roche) and HeLa cells using Lipofectamine (Invitrogen) in 24-well plates was performed according to the suppliers' instructions. Cells were harvested 40 h posttransfection and resuspended in 100 μl of Bright-Glo lysis buffer (Promega), and relative luciferase activity was determined according to Promega's instructions. The AalDNV P7/7.4 and P60 promoters were functional both in insect cells and in human cells, in contrast to constructs lacking a promoter or with promoters in the reverse orientation, whereas the SV40 promoter preferred HeLa cells and the MlDNV promoter preferred insect cells (Fig. 2C).

(i) Promoter 7/7.4 elements. The RACE experiments and RPA (Fig. 2) revealed the location of Inr1 and Inr2 at either side of the NS1 protein initiation codon (Fig. 3C). Consequently, the NS1 transcript could potentially be translated into both NS1 and NS2 proteins, whereas the NS2 transcript could only code for the NS2 protein. Different constructs, leaving Inr1 and its upstream elements intact (for NS1 transcription) but with mutations in the NS2 promoter and NS1/2 translation elements, were made with a luciferase reporter gene (Fig. 3B and C). Constructs containing the intact luciferase AUG (F1 to F7) served as a positive control for transcription, whereas those lacking the luciferase AUG (F1− to F7−) showed the impact of the various mutations.

P7/7.4 consisted of three segments (Fig. 3B). The KpnI-SacI segment I contained upstream promoter elements, the SacI-PstI segment II included both TATA boxes, ATG_{NS1}, and both Inrs and, finally, the PstI-HindIII segment III (about 20 nt for the NS1 ORF and about 70 nt for the NS2 ORF) contained the downstream promoter elements (DPE) and ATG_{NS2}. These segments were connected in pBluescript (PCR with primers containing appropriate restriction sites and pCR2.1-AalDNV template) and were transferred as a whole (or after mutation) to the pGL4.20 luciferase reporter. The diagram in Fig. 3C delineates the different constructs and knockouts (using the Transformer kit according to Clontech's instructions) (Fig. 3C). Certain ATGs were mutated to TTG, Inr (CATG) was mutated to GCCG, and the TATA box of NS2 was mutated to GCTCGAG. In addition, alternates of F1 to F7 that lacked the luciferase initiation codon (F1− to F7−) as well as two constructs from which the NS2 core promoter was mutated (F2* and F5−*) were obtained. C6/36 and 293T (as for HeLa) cells were transfected, and luciferase activity was determined. Except for F2* and F5−*, all constructs were expected to yield both NS1 and NS2 mRNAs. The observed luciferase activity in C6/36 and 293T cells (Fig. 3E) corresponded well with the expected luciferase activities, summarized in Fig. 3D, from (i) fusion proteins with NS1 (F1 and F1−) and NS2 (F5 and F5−), (ii) directly from the luciferase

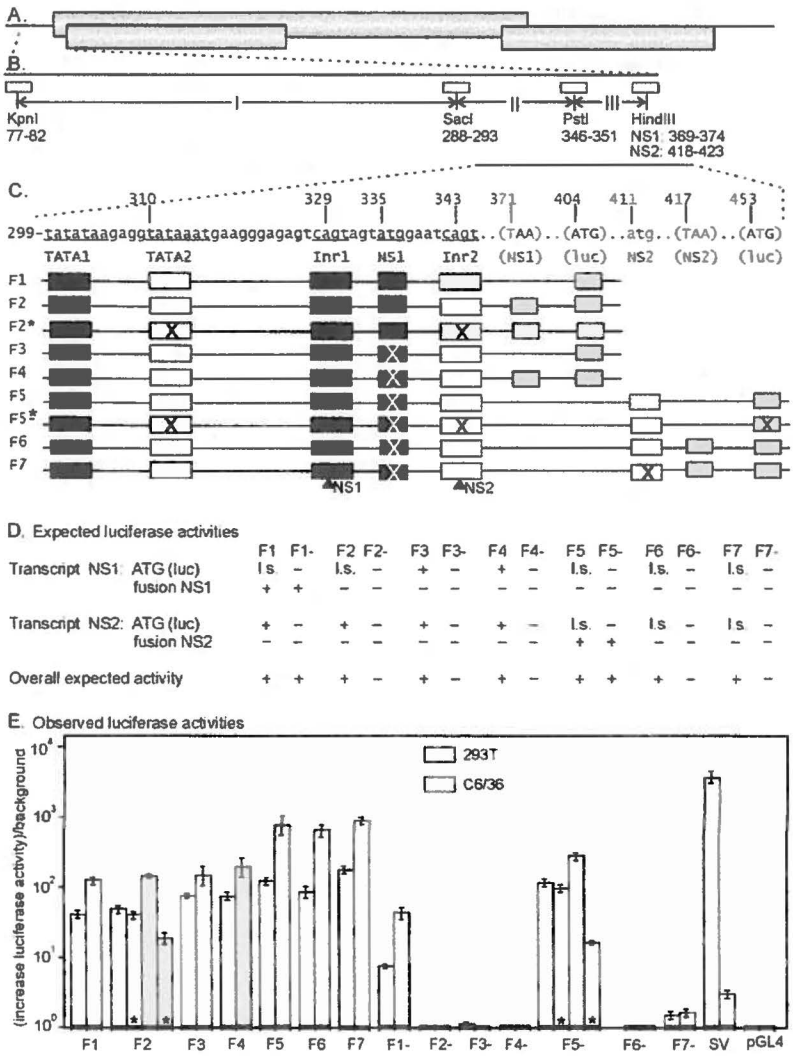


FIG 3 Analysis of AalDNV P7/7.4 promoter elements using a luciferase reporter gene. (A) Diagram of ORFs in AalDNV, where nt 77 to 423 contain typical sequences of core promoter elements for NS1 and NS2. (B) Three segments (I to III) were connected via pBluescript in pGL4 and could be individually mutated and swapped. (C) In the constructs, filled boxes represent elements thought to direct NS1 mRNA expression and open boxes those thought to direct NS2 mRNA expression, filled triangles represent transcript starts that were established, crosses represent knockouts, and gray boxes represent replaced sequences. Additionally, mutants were made for all constructs in which the initiation codon of luciferase was knocked out (F1− to F7−), and in 2 constructs TATA2 and Inr2 (for NS2 transcripts) were knocked out (indicated by *). (D) Expected activities using the luciferase reporter gene in the pGL4 vector. ("l.s." represents luciferase activity if leaky scanning occurs.) (E) The observed luciferase activity matched the expected activity, except for the very low F7− activity (none expected). These results indicated that the NS initiation codons, and hence the two sets of promoter elements for transcription, were individually important for expression but nevertheless leaky scanning on the NS1 transcript could also contribute to the expression of NS2.

initiation codon (F1 to F7), or (iii) after leaky scanning (for F1, F2, and F5 to F7). Therefore, individual elements of both NS1 and NS2 promoters contributed to mRNA expression.

Nevertheless, NS2 could be generated on the NS1 transcript by leaky scanning. Blocking NS2 transcription by mutating its core promoter in F5−* yielded an ~90% drop in luciferase activity in C6/36 cells compared to F5−. The remaining activity could be ascribed to a fusion protein from ATG_{NS2} on the NS1 transcript since the introduction of an NS2 stop codon in frame after the NS2 initiation codon in F6− or knocking out ATG_{NS2} in F7− abolished luciferase activity. Consequently, for F2* the remaining luciferase activity was due mainly to leaky scanning on the NS1 transcript and initiation from ATG_{luc}. In conclusion, although NS1 and NS2 mRNAs have their own promoter elements in p7/7.4, NS1 mRNA contributes significantly through leaky scanning to NS2 expression.

In conclusion, Northern blotting, RACE, amplicon mapping, and RPA results were all in agreement. 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}. 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 (16, 17). 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 (18), to regulate NS and VP expression.

ACKNOWLEDGMENTS

This work was supported by a grant from the Natural Sciences and Engineering Research Council of Canada to P.T. and scholarships from the Agence Universitaire de la Francophonie and the Fondation Armand-Frappier to H.T.P.

REFERENCES

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.
2. 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, United Kingdom.
3. Bergoin M, Tijssen P. 2010. Densoviruses: a highly diverse group of arthropod parvoviruses, p 57–90. *In* Asgari S, Johnson KN (ed), Insect virology. Horizon Scientific Press, Norwich, United Kingdom.
4. Boublik Y, Jousset FX, Bergoin M. 1994. Complete nucleotide sequence and genomic organization of the Aedes albopictus parvovirus (AaPV) pathogenic for Aedes aegypti larvae. Virology 200:752–763.
5. Afanasiev BN, Galyov EE, Buchatsky LP, Kozlov YV. 1991. Nucleotide sequence and genomic organization of Aedes densonucleosis virus. Virology 185:323–336.
6. Ren X, Hoiczyk E, Rasgon JL. 2008. Viral paratransgenesis in the malaria vector Anopheles gambiae. PLoS Pathog. 4:e1000135. doi:10.1371/journal.ppat.1000135.
7. Jousset FX, Barreau C, Boublik Y, Cornet M. 1993. A parvo-like virus persistently infecting a C6/36 clone of Aedes albopictus mosquito cell line and pathogenic for Aedes aegypti larvae. Virus Res. 29:99–114.
8. Barreau C, Jousset FX, Bergoin M. 1996. Pathogenicity of the Aedes albopictus parvovirus (AaPV), a denso-like virus, for Aedes aegypti mosquitoes. J. Invertebr. Pathol. 68:299–309.
9. Barreau C, Jousset FX, Bergoin M. 1997. Venereal and vertical transmission of the Aedes albopictus parvovirus in Aedes aegypti mosquitoes. Am. J. Trop. Med. Hyg. 57:4364–4370.
10. Ward TW, Kimmick MW, Afanasiev BN, Carlson JO. 2001. Characterization of the structural gene promoter of Aedes aegypti densovirus. J. Virol. 75:1325–1331.
11. Kimmick MW, Afanasiev BN, Beaty BJ, Carlson JO. 1998. Gene expression and regulation from the p7 promoter of Aedes densonucleosis virus. J. Virol. 72:4364–4370.
12. Afanasiev BN, Kozlov YV, Carlson JO, Beaty BJ. 1994. Densovirus of Aedes aegypti as an expression vector in mosquito cells. Exp. Parasitol. 79:322–339.
13. Tijssen P, Li Y, El-Far M, Szelei J, Letarte M, Zadori Z. 2003. Organization and expression strategy of the ambisense genome of densonucleosis virus of Galleria mellonella. J. Virol. 77:10357–10365.
14. 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.
15. Fediere G, El-Far M, Li Y, Bergoin M, Tijssen P. 2004. Expression strategy of densonucleosis virus from Mythimna loreyi. Virology 320:181–189.
16. Kutach AK, Kadonaga JT. 2000. The downstream promoter element DPE appears to be as widely used as the TATA box in Drosophila core promoters. Mol. Cell. Biol. 20:4754–4764.
17. Burke TW, Willy PJ, Kutach AK, Butler JE, Kadonaga JT. 1998. The DPE, a conserved downstream core promoter element that is functionally analogous to the TATA box. Cold Spring Harbor Symp. Quant. Biol. 63:75–82.
18. Hsu JY, Juven-Gershon T, Marr MT, II, Wright KJ, Tjian R, Kadonaga JT. 2008. TBP, Mot1, and NC2 establish a regulatory circuit that controls DPE-dependent versus TATA-dependent transcription. Genes Dev. 22:2353–2358.

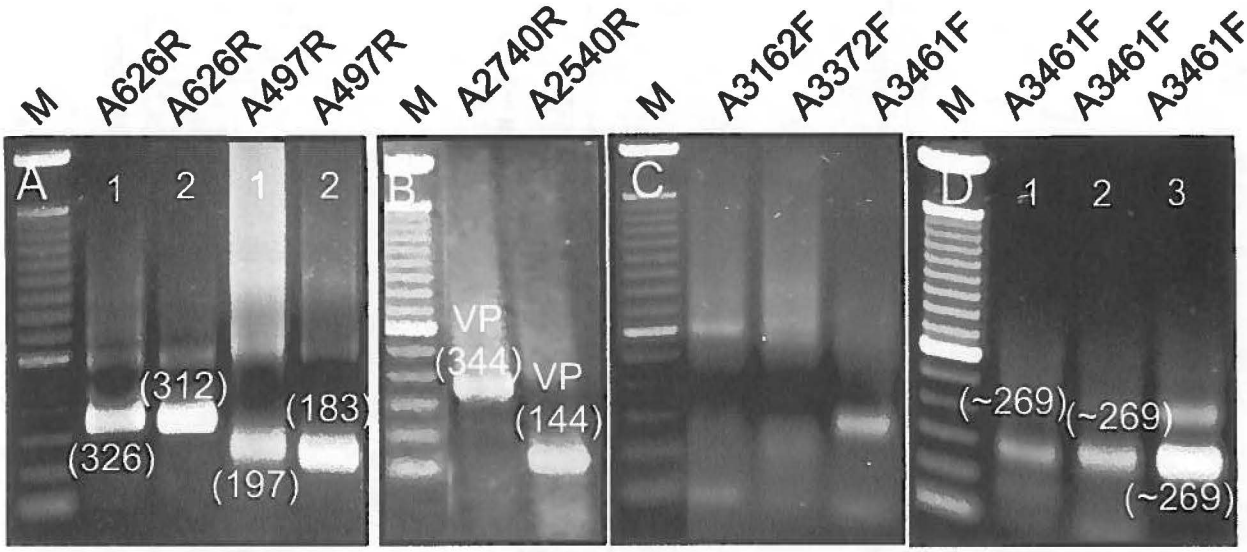


Figure S1. (A). Analysis of the amplicons obtained with different AalDNV primers and adapter primers (provided in FirstChoice™ RLM RACE kit) (M is 100 nt- size marker). 5'-RACE of AalDNV transcripts, either without digestion of the cDNA (lanes 1) or after digestion (lanes 2) with TfiI restriction enzyme gave bands at about 325 and 200 bp for NS and, (B). at about 350 and150 bp for VP.
(C). 3'RACE-PCR to identify the ends of AalDNV transcripts: lanes after a 1st round PCR with primers marked above the lanes 1, 2, 3 and D. a 2nd round (semi)-nested PCR of products from C1-C3 with the AP primer and A3461F yielded bands with sizes close to 300 nts. The precise size of the bands, as calculated after sequencing, is indicated with the respective bands.

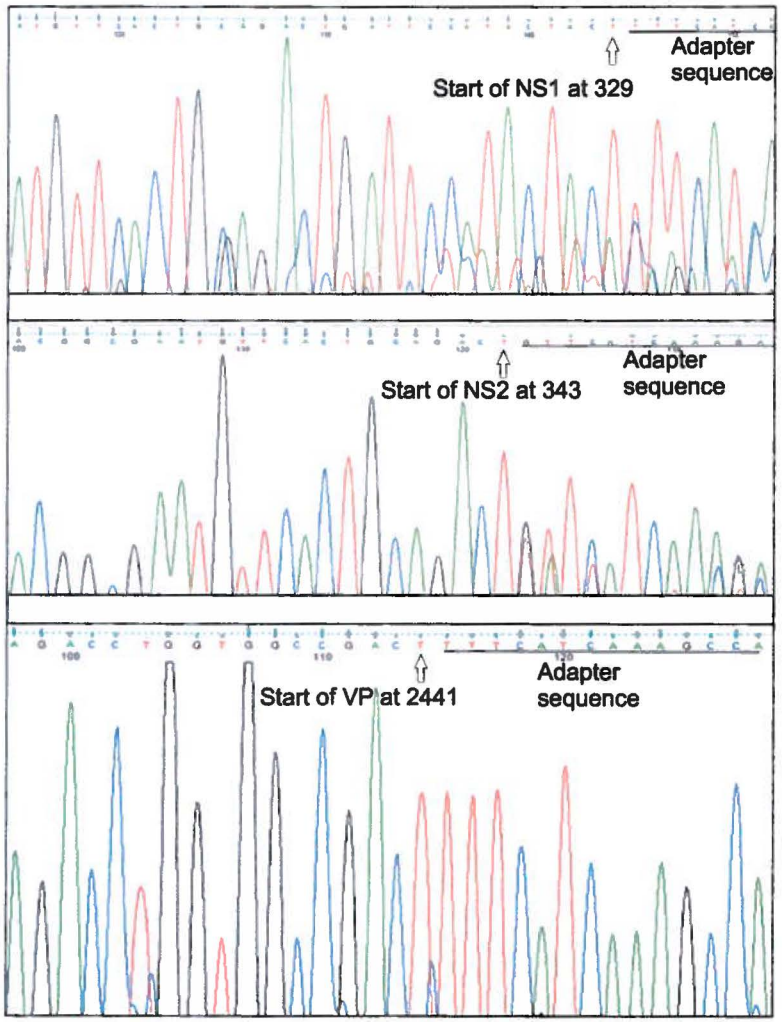


Figure S2. Starts of transcripts of NS1, NS2 and VP of AalDNV

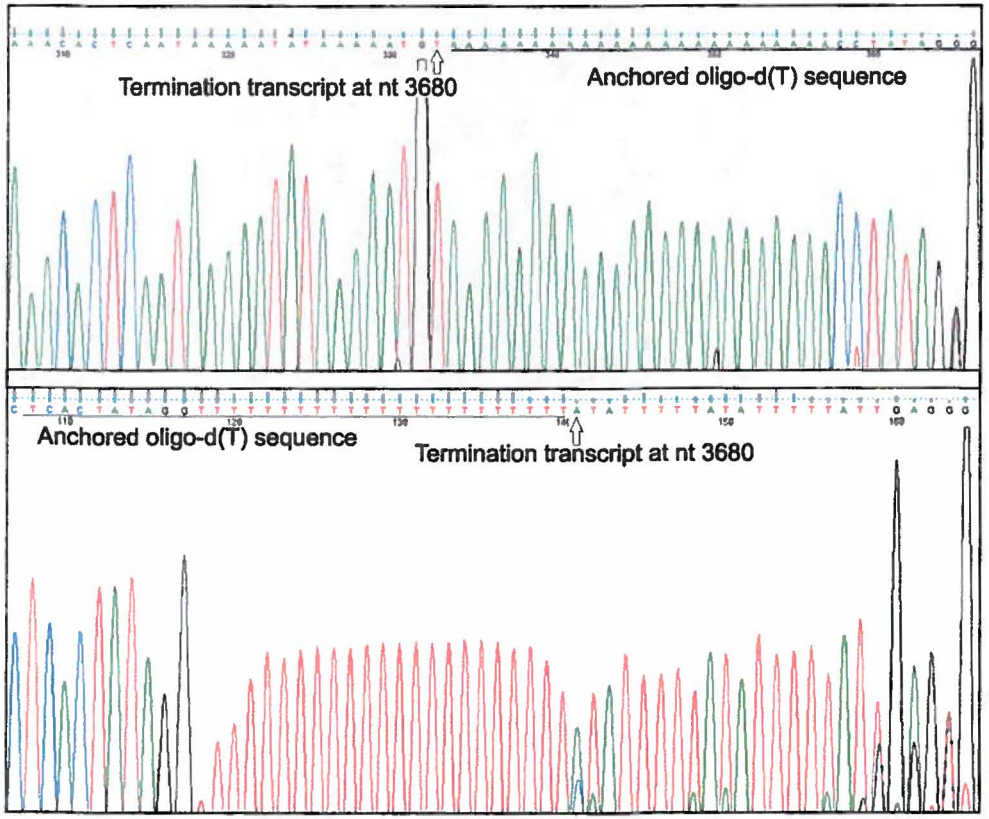


Figure S3. Sequence of termination of AalDNV transcripts.

Annexe 2. Publication: The *Acheta domesticus* densovirus, isolated from the European house cricket, has evolved an expression strategy unique among parvoviruses

JOURNAL OF VIROLOGY, Oct. 2011, p. 10069–10078 Vol. 85, No. 19
0022-538X/11/$12.00 doi:10.1128/JVI.00625-11

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 Zadori,[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, *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 (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 *Pefudensovirus* 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 70°C. The sample was vortexed after adding 125 μl isopropanol, and the DNA was then purified on High-Pure 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 (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 ^{32}P-labeled probes. The probes corresponded to a 1.5-kb BglII-SalI 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 ProVP1F and ProVP1R primers were used for the VP1 promoter, and the PrNSMf and PrNSMr primers were used for the *Mythimna loreyi* densovirus (MlDNV) 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 Ad1HAR, 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	TTTGTTGCAATCCCATAATAGTAC	2610–2633	Near 3′ end of VP mRNA
NAdR	gctctagatCATCTTGAACGTTTACCACCACT	3892–3915	Just upstream of VP4
Adsp	tcggaattcCACGTTCTTGTGGATGAGG	4362–4380	19-37 nt into VP2
Adsvp	gccTACCAGAAATCCGTGTAATGACA	4546–4534/4403–4393	Small intron splice
Ad3s	CGTGAGTACTGATACTTTTTATTT	4435–4412	End of ORF-B (TGA)
Ads	gccCCTCAACAACCTAAAAAACGTGAGTACTGA	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	gccCCTGTATATCCTCAACAACCTAAA	4462–4439	End of ORF-B (−4 aa)
AdlgF	tcggaattcATGTCTGGCGTCTTTACA	5230–5213	Start of VP1 (ORF-B)
RNAtag	gggtctagagctcgagT_{17}	Poly(A) sequence	3′/5′-RACE (first round)
ADAP	gggtctagagctcgagT		Subsequent rounds of RACE
ProNSF	acggtaccGATATAAAGAGCAAGCACCC	109–128	NS1 promoter (KpnI)
ProNSR	gaagatCTGTGCTGGAGGCGCTTCTACTGCAGCGAACAAC GTACCTGAGTTCCAGAACAC	240–207	NS1 promoter (reverse)
ProVP1R	gaagatctAAGACGCCAGAGATTTAATACT	5217–5238	VP1 (BglII site; reverse)
ProVP1F	gtaggtaccGATATAAAGAGCAAGCACCCA	5323–5300	VP1(KpnI site; forward)
PrNSMf	ACGGTACCGACTATAAATAGAGCTGAGC		MlDNV (forward)
PrNSMr	GAAGATCTATCTTGCAATAGATATACCTA		MlDNV (reverse)
pFECRIF	CGCAAATAAATAAGAATTCTACTGTTTTCGTAAC		Mutation in pFastBac1
pFECRIR	GTTACGAAAACAGTAGAATTCTTATTTATTTGCG		Mutation in pFastBac1
Ad1HAR	GCTCTAGATCAAGCGTAATCTGGAACATCGTATGGGTA TTTTTGTTGCAATCCCATAATA		VP products (reverse primer [HA])
AdATG1B	ggaattcATGTCCGTCTTTACAGATCTCAC		
AdmATG1B	ggaattcACCTCCGTCTTTACAGATCTCTCAC		

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

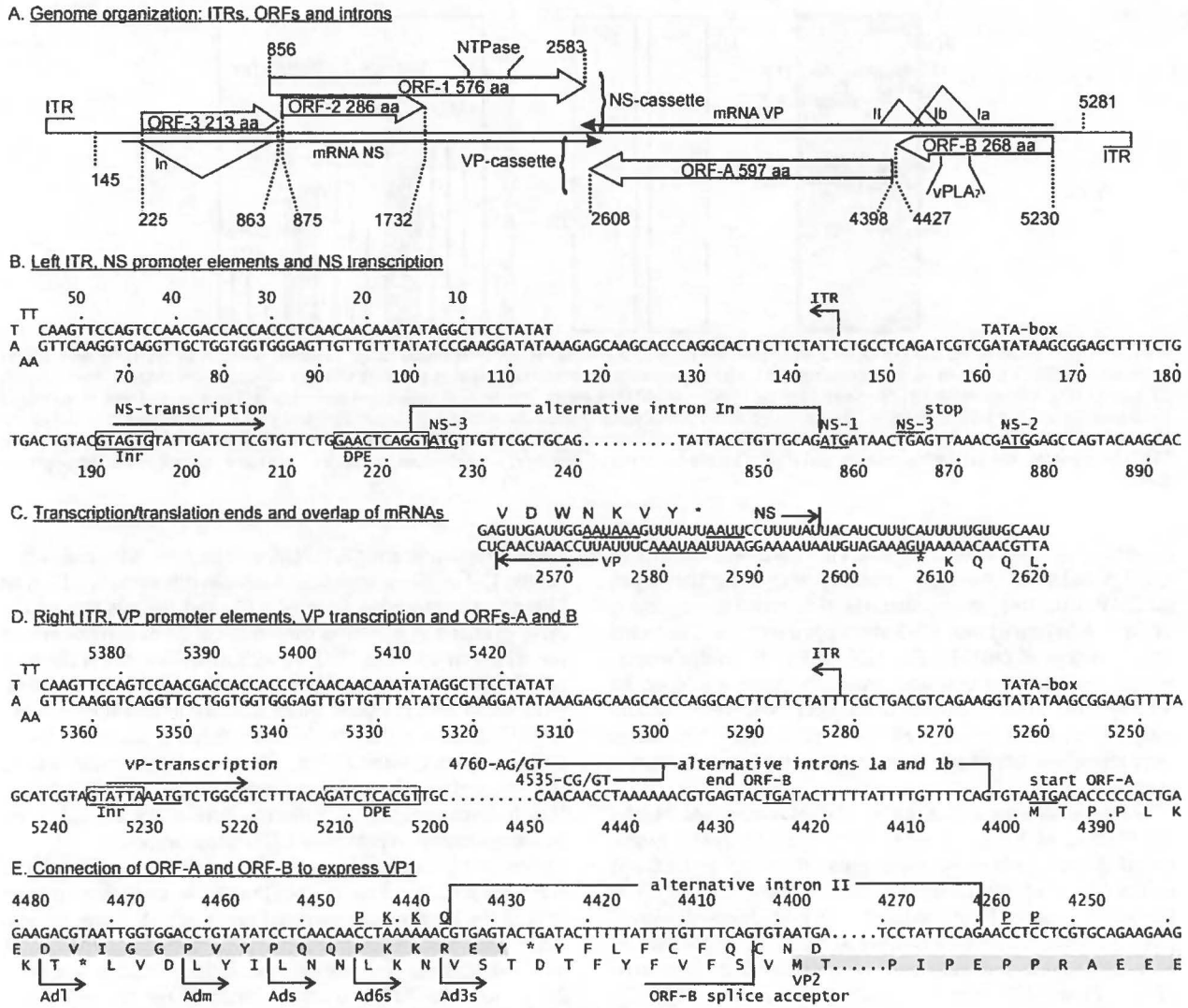


FIG. 1. Genome organization of AdDNV. (A) Overview of genome organization, positions, and sizes of ITRs, ORFs, and introns. The In intron in the NS mRNA, between nt 223 and nt 855, occurs in about half of the NS transcripts. Ia (nt 4403 to 4758), Ib (nt 4403 to 4533), and II (nt 4260 to 4434) are introns 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 *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.

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

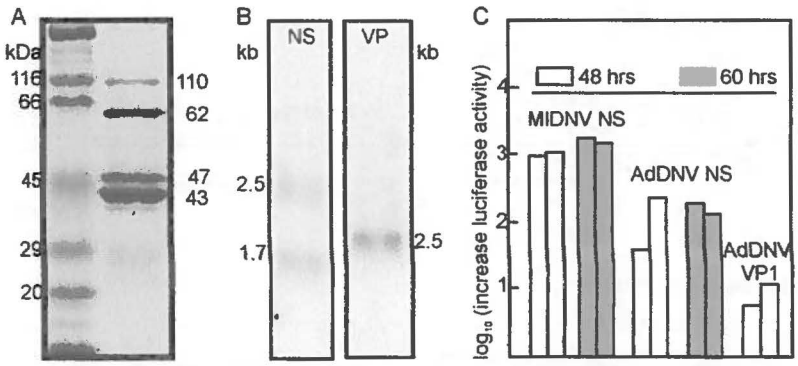


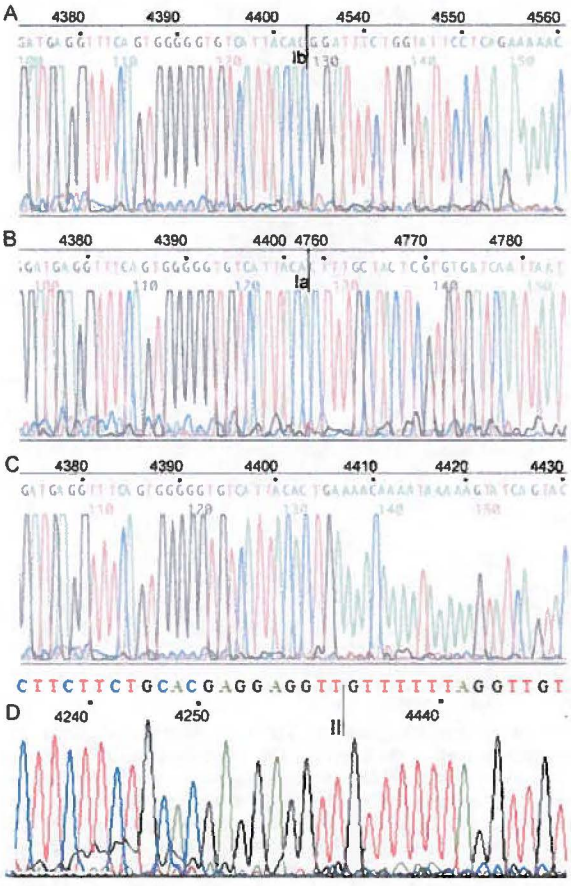
FIG. 2. (A) SDS-PAGE analysis of structural proteins of AdDNV. Lane 1, Bio-Rad broad-range standard proteins as markers; lane 2, four proteins of AdDNV (VP1 to -4, in decreasing size). The estimated masses corresponded reasonably with the sequence-predicted masses (88, 65, 51, and 47 kDa, respectively). (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 MlDNV, 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-



E Mass spectrometry of capsid proteins

VP1

```
  1 MSGVFTDLTL RDFSLGTNNL PLANTPFLRN RFGRGWNRSH PYDRLPTNEP EPLRETSFQE
 61 AAGPEETRID IAEDEINAGE GAAQAETSFS TGVEETALEL GEAATETTGL LGGATAGSAA
121 AGTAGTLGTV AATAAGGAAL AGIGIGIKKL IDHTSSKGAV LPGTDFVGPG NPIDFKPARS
181 ETDQIAKEHD LGYEDLLHRA KSQYFTEEDF KTEVYKLDDE AIHRFSEEYQ KSGTMQAFVG
241 KTGLKAKRLI EDVIGGPVTP QQPKKQAPRA EEEEVAPATP TADDILDELP EGNPEDIDDF
301 DVDQLPSTTQ EHHPTTDLAA QPGRSTDPDP EPVSNPIDST PPDEGDSSSM SAPLTPPITP
361 DKTQDKGKGK RGGGGRPPKS SGGKRSRMAG TSLPGAGADI AGEGGGQAII PIPRPITTPH
421 NNFIIFNKVH RFLTYGLANV VIPITRTVND VTYVDNFVIT SLGRLPVDRP FLYMNPNEFA
481 QLPPGSCIEE CNVRVTAFTP RIAFQTNSSN TGLATLNQNQ FILHATGLNI KTQGVDVRPK
541 EFQANEPMVV SSIDELGAIR DENLFEDYVK EFYGDKSVPN HVPRHQPGIP YPLKNYFALV
601 LQQNGTADNP GYECLQSHVN EIRCNGPPGE IVEYNYKPQL GLIKKAIPAV YTGVPSAVGG
661 NSTLSCPAGA GNSQYRTADF TMTNNILRGD SETFKDTVVS NTEWTIATQI EKSQILHAGA
721 FPSYTPPAQP SLHIGVKPVH ALTTANLDDN LNNANFTDTQ AYFIVEAACK VRLQYPSIRP
781 LFNGPNTIPD NQIYSSSTLQ YTHAVGKSTI MGLQQK
```

VP2

```
  1 MTPPLKPHPQ ERDNWEYLNE GQRRYAVEQW QLARVRRGLP IDHPIPEPPR AEEEEVAPAT
 61 PTADDILDEL PEGNPEDIDD FDVDQLPSTT QEHHPTTDLA AQPGRSTDPD PEPVSNPIDS
121 TPPDEGDSSS MSAPLTPPIT PDKTQDKGKG KRGGGGRPPK SSGGKRSRMA GTSLPGAGAD
181 IAGEGGGQAI IPIPRPITTP HNNFIIFNKV HRFLTYGLAN VVIPITRTVN DVTYVDNFVI
241 TSLGRLPVDR PFLYMNPNEF AQLPPGSCIE ECNVRVTAFT PRIAFQTNSS NTGLATLNQN
301 QFILHATGLN IKTQGVDVRP KEFQANEPMV VSSIDELGAI RDENLFEDYV KEFYGDKSVP
361 NHVPRHQPGI PYPLKNYFAL VLQQNGTADN PGYECLQSHV NEIRCNGPPG EIVEYNYKPQ
421 LGLIKKAIPA VYTGVPSAVG GNSTLSCPAG AGNSQYRTAD FTMTNNILRG DSETFKDTVV
481 SNTEWTIATQ IEKSQILHAG AFPSYTPPAQ PSLHIGVKPV HALTTANLDD NLNNANFTDT
541 QAYFIVEAAC KVRLQYPSIR PLFNGPNTIP DNQIYSSSTL QYTHAVGKST IMGLQQK
```

VP3

```
  1 MSAPLTPPIT PDKTQDKGKG KRGGGGRPPK SSGGKRSRMA GTSLPGAGAD IAGEGGGQAI
 61 IPIPRPITTP HNNFIIFNKV HRFLTYGLAN VVIPITRTVN DVTYVDNFVI TSLGRLPVDR
121 PFLYMNPNEF AQLPPGSCIE ECNVRVTAFT PRIAFQTNSS NTGLATLNQN QFILHATGLN
181 IKTQGVDVRP KEFQANEPMV VSSIDELGAI RDENLFEDYV KEFYGDKSVP NHVPRHQPGI
241 PYPLKNYFAL VLQQNGTADN PGYECLQSHV NEIRCNGPPG EIVEYNYKPQ LGLIKKAIPA
301 VYTGVPSAVG GNSTLSCPAG AGNSQYRTAD FTMTNNILRG DSETFKDTVV SNTEWTIATQ
361 IEKSQILHAG AFPSYTPPAQ PSLHIGVKPV HALTTANLDD NLNNANFTDT QAYFIVEAAC
421 KVRLQYPSIR PLFNGPNTIP DNQIYSSSTL QYTHAVGKST IMGLQQK
```

VP4

```
  1 MAGTSLPGAG ADIAGEGGGQ AIIPIPRPIT TPHNNFIIFN KVHRFLTYGL ANVVIPITRT
 61 VNDVTTVDNF VITSLGRLPV DRPFLYMNPN EFAQLPPGSC IEECNVRVTA FTPRIAFQTN
121 SSNTGLATLN QNQFILHATG LNIKTQGVDV RPKEFQANEP MVVSSIDELG AIRDENLFED
181 YVKEFYGDKS VPNHVPRHQP GIPYPLKNYF ALVLQQNGTA DNPGYECLQS HVNEIRCNGP
241 PGEIVEYNYK PQLGLIKKAI PAVYTGVPSA VGGNSTLSCP AGAGNSQYRT ADFTMTNNIL
301 RGDSETFKDT VVSNTEWTIA TQIEKSQILH AGAFPSYTPR AQPSLHIGVK PVHALTTANL
361 DDNLNNANFT DTQAYFIVEA ACKVRLQYPS IRPLFNGPNT IPDNQIYSSS TLQYTHAVGK
421 STIMGLQQK
```

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 (His_6-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 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, AdIgF/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-

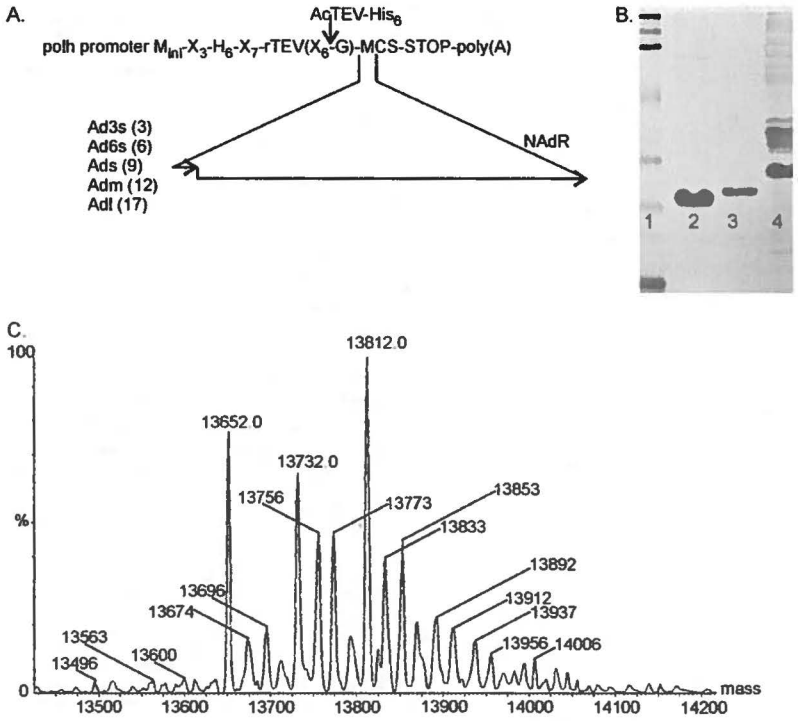


FIG. 4. (A) Construction of various ORF-A and ORF-B constructs with different Ad primers (Ad3s, Ad6s, etc.) linking a short part of ORF-B to a larger part of ORF-A, using the NAdR primer and the pFastbacHTb baculovirus system. Upstream of the insert is a TEV proteolytic site and a histidine tail (H6); 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, 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.

tiation 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 MlDNV, 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 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 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

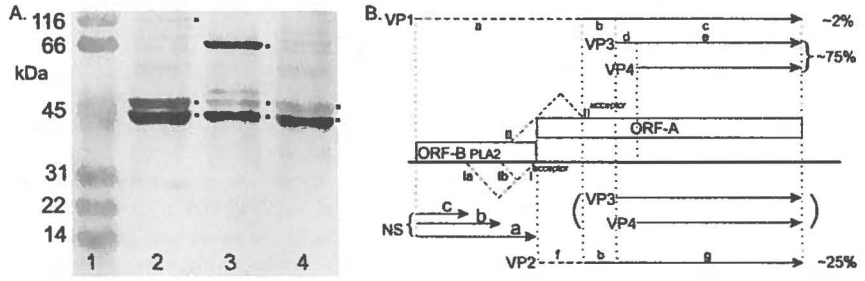


FIG. 5. (A) Western blot of pFastbac VP constructs. Lane 1, prestained SDS-PAGE protein standards (Bio-Rad). Lane 2, expression products from the ORF-A–ORF-B construct from which intron II was deleted. Strong bands were observed for VP3 and VP4, with a weak band for VP1 (all indicated with dots), in 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,

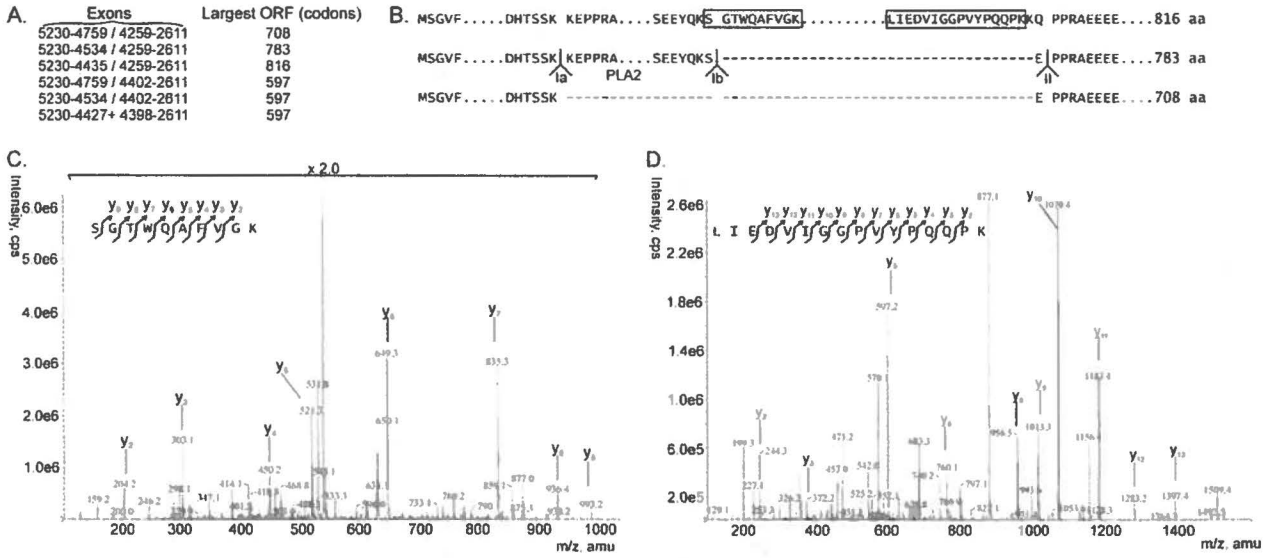


FIG. 6. Potential VP1 proteins encoded by the VP gene cassette. (A) Different potential splice combinations of the ORFs in the VP gene cassette and the largest protein that could then be coded. N-terminal amino acid sequencing showed that the 597-codon ORF corresponded to VP2 and, theoretically, that 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 SGTWQAFVGK 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 Fig. S2 in the supplemental material.

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

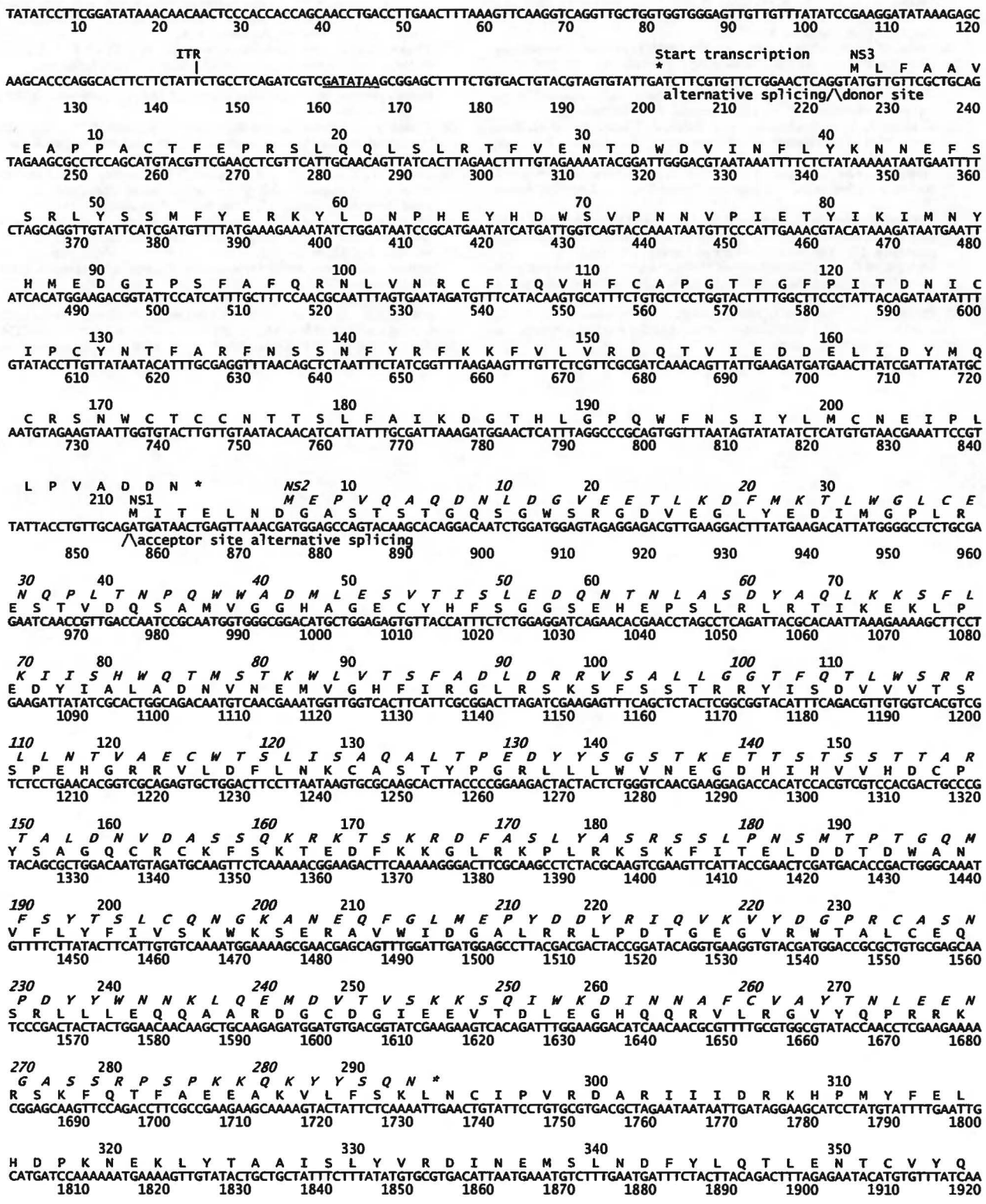
P.T. acknowledges support from the Natural Sciences and Engineering Research Council of Canada.

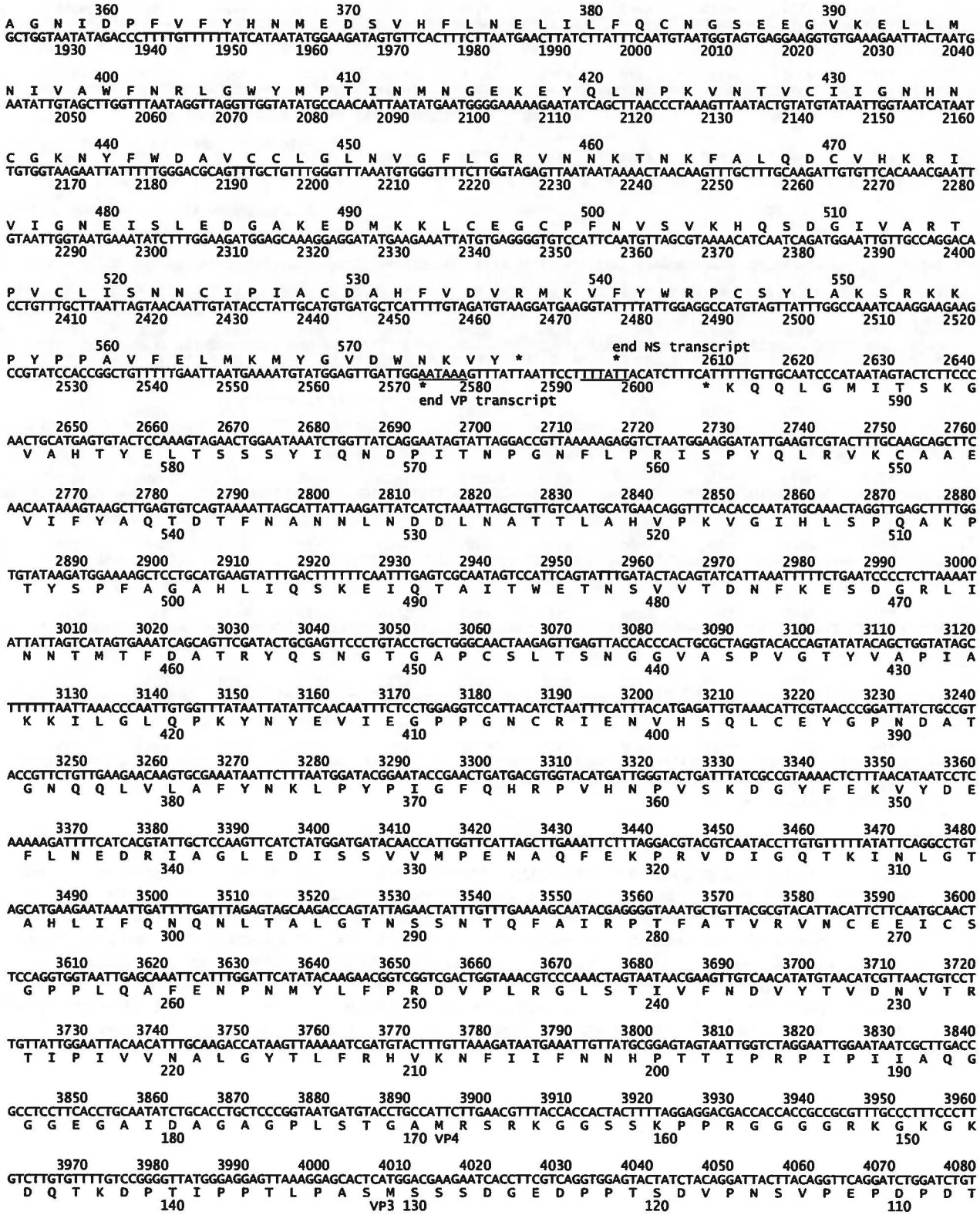
We thank the professionals in the cricket-rearing industry for their work.

REFERENCES

1. **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.
2. **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.
3. **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.
4. **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.
5. **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.
6. **Fediere, G., Y. Li, Z. Zadori, J. Szelei, and P. Tijssen.** 2002. Genome organization of Casphalia extranea densovirus, a new iteravirus. Virology 292:299–308.
7. **Fernandes, S., and P. Tijssen.** 2009. Seamless cloning and domain swapping of synthetic and complex DNA. Anal. Biochem. 385:171–173.
8. **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.
9. **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.
10. **Kaufmann, B., et al.** 2010. Structure of Penaeus stylirostris densovirus, a shrimp pathogen. J. Virol. 84:11289–11296.
11. **Kaufmann, B., et al.** 2011. Structure of Bombyx mori densovirus 1, a silkworm pathogen. J. Virol. 85:4691–4697.
12. **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.
13. **Laemmli, U. K.** 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature 227:680–685.
14. **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.
15. **Matsudaira, P.** 1987. Sequence from picomole quantities of proteins electroblotted onto polyvinylidene difluoride membranes. J. Biol. Chem. 262: 10035–10038.
16. **Meynardier, G., et al.** 1977. Virus de type densonucléose chez les orthoptères. Ann. Soc. Entomol. Fr. 13:487–493.
17. **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 A 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. **Szelei, 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 (paravovirus). 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.


ITR
Start transcription
NS3
alternative splicing/\donor site
NS1
/\acceptor site alternative splicing


end NS transcript
end VP transcript
VP4
VP3

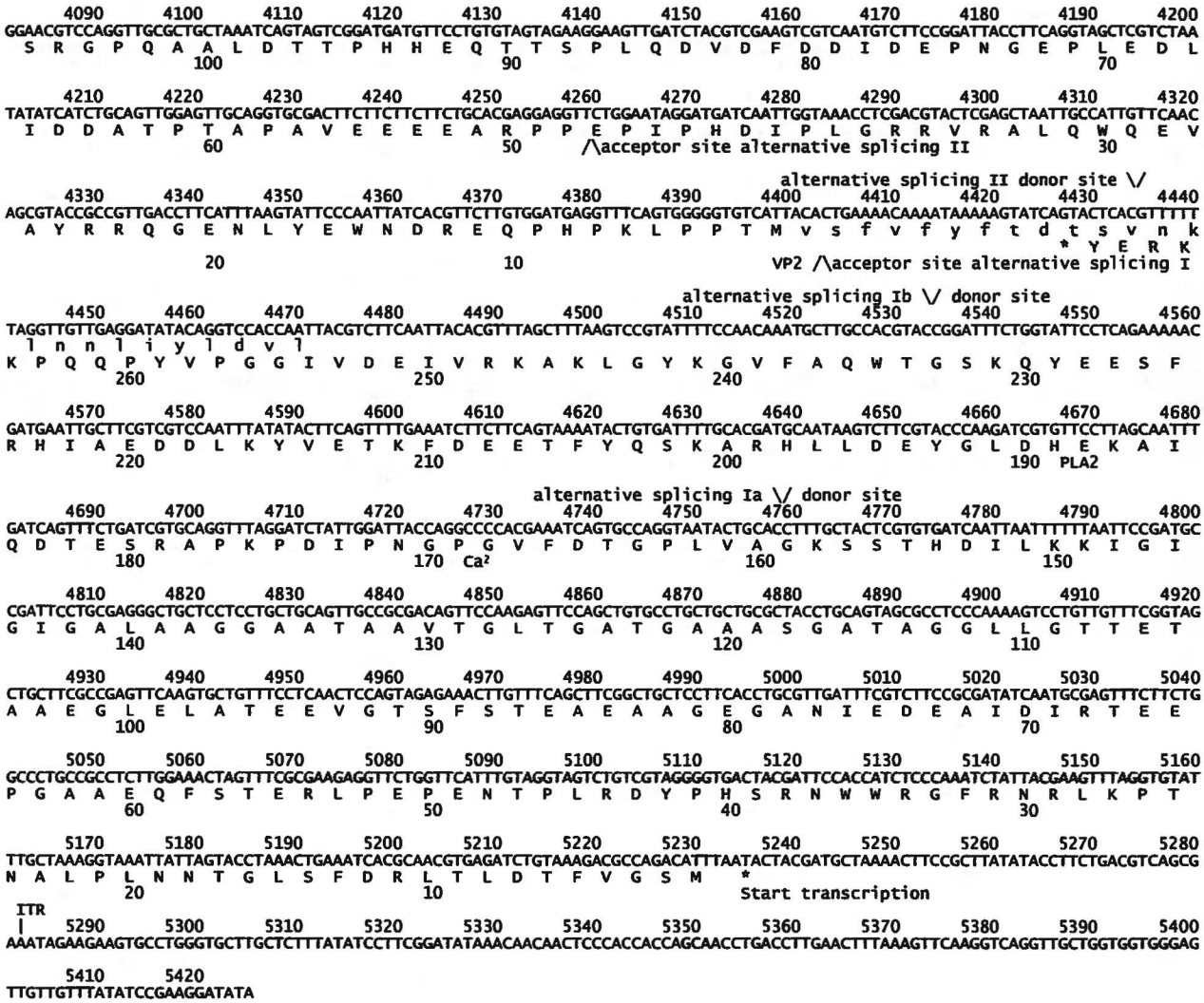


Figure S1. Annotated sequence of the AdDNV strain isolated in 1977 by Meynardier et al. Translated sequence above nucleotide sequence is from sense mRNA, whereas the translation below the translation is from the antisense mRNA. Both the nucleic acid and the proteins are numbered; however, the NS2 sequence 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 Ca^2 (Ca^2) 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. 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).

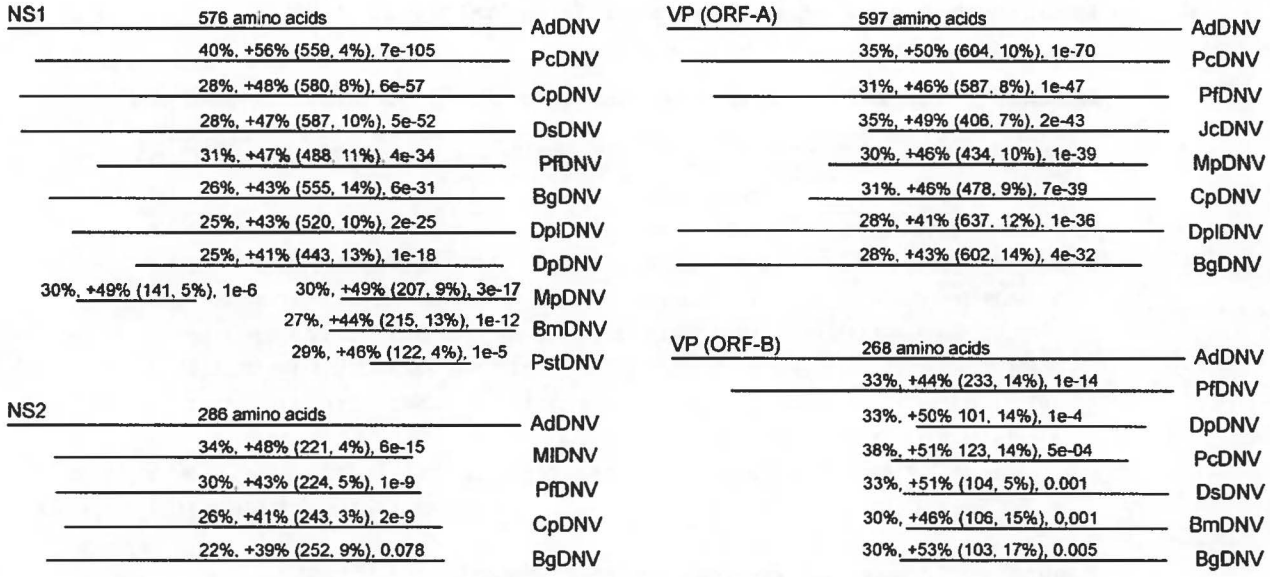


Figure S3. Similarities of AdDNV ORF sequences with those of other densoviruses. The lines represent the location of the homologous domains in the other densoviruses. Above each line, the identity %, positives %, (length of domain in amino acids, gap %), likelihood of similarity by chance, is noted. 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).

Annexe 3. Publication: Comparative genomic analysis of *Acheta domesticus* densovirus isolates from different outbreaks in Europe, North America, and Japan

genomeA
Journals.ASM.org

Comparative Genomic Analysis of *Acheta domesticus* Densovirus Isolates from Different Outbreaks in Europe, North America, and Japan

Hanh T. Pham,[a] Hajime Iwao,[b] Jozsef Szelei,[a] Yi Li,[c] Kaiyu Liu,[d] Max Bergoin,[a] Peter Tijssen[a]

INRS, Institut Armand-Frappier, Laval, QC, Canada[a]; Marinpia Nihonkai 5932-445, Nishifunami, Niigata, Japan[b]; Bioengineering Department, Wuhan Bioengineering Institute, Yangluo, Wuhan, Hubei, People's Republic of China[c]; College of Life Sciences, Central China Normal University, Wuhan, Hubei, People's Republic of China[d]

The first densovirus from a cricket, *Acheta domesticus* densovirus (AdDNV) (*Parvoviridae*), was isolated in Europe in 1977 and has been studied previously. We compared seven additional AdDNV genomes isolated from 4 other European outbreaks, 2 major North American outbreaks, and a Japanese outbreak. Phylogenetic analysis suggested that the 2009 Japanese and North American outbreaks were not related.

Received 15 July 2013 **Accepted** 22 July 2013 **Published** 15 August 2013

Citation Pham HT, Iwao H, Szelei J, Li Y, Liu K, Bergoin M, Tijssen P. 2013. Comparative genomic analysis of *Acheta domesticus* densovirus isolates from different outbreaks in Europe, North America, and Japan. Genome Announc. 1(4):e00629-13. doi:10.1128/genomeA.00629-13.

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

The United States cricket pet food industry supplies billions of live crickets (*Acheta domesticus* L.) per year to pet stores and individuals. This industry has been devastated by epizootic *Acheta domesticus* densovirus (AdDNV) outbreaks since 2009 (1). Alternative, presumably virus-resistant, field cricket species have been introduced and led to widespread United States (and European) distribution of exotic *Gryllus* species, such as the naturally widespread African, European, and Asian "black cricket," *G. bimaculatus*, and the previously unknown red cricket, *G. locorojo*, despite existing United States federal regulations to prevent such movement (2).

Previously, AdDNV outbreaks have occurred in Europe, and a small, contained one occurred in an operation in the Southeastern United States (3). The isolate from this U.S. outbreak, preserved in paraffin, was sequenced and was found to be identical to the European isolate from 1977 (1). The 1977 isolate (4) was cloned and sequenced and its expression strategy studied (5). In particular, expression of the structural proteins was found to differ from those of other parvoviruses, since two large open reading frames (ORFs) were spliced and VP2 was not completely contained within VP1, i.e., structural proteins did not form a nested set of N-extended proteins like those of other densoviruses. The *Blattella germanica* densovirus (BgDNV) was found to have adapted a similar strategy (6).

AdDNV samples were obtained from different outbreaks: June 2004 (AdEu04), July 2006 (AdEu06), May 2007 (AdEu07), and August 2009 (AdEu09), all in Germany; September 2009 (AdNA09) in the United States and Canada; and December 2012/January 2013 in Japan (AdJP12). Moreover, AdDNV was isolated from different cricket species, *Gryllus sigillatus* in September 2012 in Canada (GsCa12) and *Gryllus locorojo* in the United States in March 2012 (GlUS12). For all isolates, DNA samples were isolated as previously described (5). Primers (5) were designed on the sequence obtained previously for the 1977 AdDNV isolate from Europe (GenBank accession number HQ827781), so that overlapping amplicons could be obtained, whereas termini were cloned as described previously (7, 8).

At least two complete clones of each isolate were sequenced in both directions, and four times at locations of discrepancies, using Sanger's method and the primer-walking method as described previously (9). Contigs were assembled by use of the CAP3 program (http://pbil.univ-lyon1.fr/cap3.php/) (10). Like some other densoviruses (11–16), AdDNV was found to have a broad host range, infecting at least *Acheta domesticus*, *Gryllus locorojo*, and *Gryllus sigillatus*. In fact, AdNA09 and the GlUS12 isolates had the same sequence.

All isolates had genomes of 5425 nucleotides (nt) and 144 nt-inverted terminal repeats, of which the 114 distal nt formed perfect hairpins. The location of open reading frames, TATA boxes, and splicing sites were all conserved compared to those of the original 1977 isolate (5). The highest protein sequence identity among the isolates was found for NS1 and its overlapping NS2 (both 99.3%), whereas NS3 had an identity of 94.4%, and the two ORFs, A and B, of the structural proteins had identities of 98.1 and 97.1%, respectively. Interestingly, phylogenetic analysis showed that the AdJP12 and GsCa12 clade diverged early from the European/North America clade, probably 20 years before the epidemics occurred simultaneously in 2009 in North America and Japan. This suggested another contributing factor to these outbreaks.

Nucleotide sequence accession numbers. The GenBank accession numbers are KF015274 for the AdDNV-AdEu04, KF015275 for the AdDNV-AdEu06, KF015276 for the AdDNV-AdEu07, KF015277 for the AdDNV-AdEu09, KF015278 for the AdDNV-AdNA09 and AdDNV-GlUS12, KF015279 for the AdDNV-AdJP12, and KF015280 for the AdDNV-GsCa12 isolates.

ACKNOWLEDGMENTS

This work was supported by the Natural Sciences and Engineering Research Council of Canada to P.T.; H.T.P. acknowledges tuition fee waivers at INRS.

REFERENCES

1. 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.
2. 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.
3. Styer EL, Hamm JJ. 1991. Report of a densovirus in a commercial cricket operation in the southeastern United-States. J. Invertebr. Pathol. 58: 283–285.
4. Meynadier G, Matz G, Veyrunes J-C, Bres N. 1977. Virus de type densonucléose chez les orthoptères. Ann. Soc. Entomol. Fr. 13:487–493.
5. 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.
6. Kapelinskaya TV, Martynova EU, Schal C, Mukha DV. 2011. Expression strategy of densonucleosis virus from the German cockroach, *Blattella germanica*. J. Virol. 85:11855–11870.
7. Huynh OT, Pham HT, Yu Q, Tijssen P. 2012. *Pseudoplusia includens* densovirus genome organization and expression strategy. J. Virol. 86: 13127–13128.
8. Pham HT, Nguyen OTH, Jousset F-X, Bergoin M, Tijssen P. 2013. *Junonia coenia* densovirus (JcDNV) genome structure. Genome Announc. 1(4):e00591-13. doi:10.1128/genomeA.00591-13.
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.
10. Huang X, Madan A. 1999. CAP3: A DNA sequence assembly program. Genome Res. 9:868–877.
11. 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.
12. Bergoin M, Tijssen P. 2010. Densoviruses: a highly diverse group of arthropod parvoviruses, p 57–90. *In* Asgari S, Johnson KN (ed), Insect virology. Horizon Scientific Press, Norwich, United Kingdom.
13. 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, United Kingdom.
14. Fédière G, El-Far M, Li Y, Bergoin M, Tijssen P. 2004. Expression strategy of densonucleosis virus from *Mythimna loreyi*. Virology 320: 181–189.
15. El-Mergawy R, Yi L, El-Sheikh M, El-Sayed M, Abol-Ela S, Bergoin M, Tijssen P, Fédière G. 2003. Epidemiology and biodiversity of the densovirus MlDNV in the field populations of *Spodoptera littoralis* and other noctuid pests. Bull. Fac. Agric. Cairo Univ. 54:269–281.
16. Fédière G, Ul-Sheikh M, El Mergawy R, Salah M, Masri M, Abol-Ela S, Bergoin M, Tijssen P. 2003. Record of densoviral diseases occurring among the noctuid populations in Lucerne fields at El-Bahareya oasis in Egypt. IOBC WPRS Bulletin 26:237–240.

Annexe 4. Publication: *Pseudoplusia includens* densovirus genome organization and expression strategy

JVI
Journals.ASM.org

GENOME ANNOUNCEMENT

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).

Densoviruses 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 pJazz 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 (MlDNV) (4). 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 (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.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 looplure-baited traps. J. Econ. Entomol. 75:647–650.
2. **Chao Y-C, Young SY III, Kim KS, Scott HA.** 1985. A newly isolated densonucleosis 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 densovirus 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 densonucleosis 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@iaf.inrs.ca.

doi: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.

Annexe 5. Publication: *Junonia coenia* densovirus (JcDNV) genome structure

genomeA
Journals.ASM.org

Junonia coenia Densovirus (JcDNV) Genome Structure

Hanh T. Pham,[a] Oanh T. H. Huynh,[a] Françolse-Xavière Jousset,[b]* Max Bergoin,[a] Peter Tijssen[a]

INRS-Institut Armand-Frappier, Laval, QC, Canada[a]; Laboratoire de Pathologie Comparée, Université Montpellier II, Montpellier, France[b]

* Retired.

The sequence of *Junonia coenia* densovirus was the first densovirus genome sequence published, but the first published sequence contained incomplete inverted terminal repeats and ambiguous nucleotides or indels leading to an incorrect map of the open reading frames. Our sequencing of clones of the complete genome demonstrated that this virus is closely related to other viruses in the *Densovirus* genus.

Received 2 July 2013 **Accepted** 8 July 2013 **Published** 8 August 2013

Citation Pham HT, Huynh OTH, Jousset F-X, Bergoin M, Tijssen P. 2013. *Junonia coenia* densovirus (JcDNV) genome structure. Genome Announc. 1(4):e00591-13. doi:10.1128/genomeA.00591-13.

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

The common buckeye (*Junonia coenia* [Hübn.]) is a butterfly in the *Nymphalidae* family. A nonoccluded, small virus was isolated from caterpillar cadavers (1) and subsequently characterized as a densovirus (2). This *Junonia coenia* densovirus (JcDNV) was kindly provided to us by T. W. Tinsley (NERC Institute of Virology, Oxford). Its restriction map showed a close relatedness to the GmDNV densovirus from *Galleria mellonella* (3). An infectious JcDNV genome was cloned (pBRJ) and sequenced (4, 5).

Related densoviruses have been isolated, cloned, and sequenced, e.g., from *Galleria mellonella* [GmDNV (6)], *Mythimna loreyi* [MlDNV (7)], *Helicoverpa armigera* [HaDNV (8)], and *Pseudoplusia includens* [PiDNV (9)]. Compared to the published genome sequences of this group, the published JcDNV sequence had a different genome orientation, incomplete inverted terminal repeats (ITRs), and a different map of open reading frames (ORFs). In addition, there are several ambiguous sequences in the reported genome sequence of JcDNV (GenBank accession number NC_004284).

Here, the entire JcDNV genome was extracted from the same virus stock as that used to produce pBRJ, recloned, and sequenced and compared to the genome sequence of pBRJ. The separately encapsidated, complementary DNA strands reannealed upon extraction. The central 5.45-kb part of the genome sequence, after BamHI digestion, was cloned into pBluescript KS. The rest of the inverted terminal repeats (ITRs) were obtained by PCR from gel-purified 2.4- and 3.6-kb dsDNA fragments after digestion with NdeI. DNA was heated up to 95°C for 5 min, and 10 ng of each fragment was used for ligation with a 5P-ACGCAAGTACCGTGGTACCATGGATCCGG-3C3 adapter, including 1% DMSO, 1 mM hexamine cobalt chloride, and 10% polyethylene glycol (PEG) (final concentrations) and T4 RNA ligase (NEB), and then incubated overnight at room temperature. The DNA was then precipitated and eluted in 70 μL of sterile water. Five microliters was used for 25 μL of PCR, including 6% DMSO, 1.3 M betaine, 50 μM 7-deaza-dGTP, CTTCGGATCCTCT CCATCATC, and CCGGATCCATGGTACCACGGTACTT-GCGT as specific and adapter primers, respectively, and Phusion High-Fidelity DNA polymerase (98°C, 3 min; 25 cycles of 98°C 10 s, 65°C 20 s, and 72°C 20 s; elongation, 72°C, 5 min). Amplicons were cloned into a pGEMT-easy TA vector (Promega) and transformed into Sure cells. Two complete clones were sequenced in both directions and pBRJ was sequenced at locations of discrepancies, using Sanger's method and the primer-walking method as described before (6). The contigs were assembled by using the CAP3 program (http://pbil.univ-lyon1.fr/cap3.php/) (10).

The 6,032-nucleotide (nt) JcDNV genome contained 547-nt-long, nearly identical ITRs, the only differences being 396G, 537G, and the two TATA boxes (539-TATAAAT for the NS-gene cassette and 5488-TATATAA for the VP-gene cassette). The typical terminal Y-shaped hairpins of 130 nt contained two orientations at sequences 51 to 80 and 5953 to 5982, "flip" and its reverse complement orientation "flop." The ambisense JcDNV genome sequence was 83 to 87% identical to those of other viruses in the *Densovirus* genus. The ORFs, in contrast to the previous entry in GenBank, were conserved with members of the *Densovirus* genus, as were the splicing sites (11) with those identified for GmDNV (6) and MlDNV (7). Also, sequencing errors, gaps, and ambiguous nts were corrected and the ITRs completed. JcDNV contained the typical NS-1 helicase superfamily III and VP phospholipase A2 (12) motifs observed in other parvoviruses.

Nucleotide sequence accession number. The GenBank accession number of JcDNV is KC883978.

ACKNOWLEDGMENTS

This work was supported by funding from the Natural Sciences and Engineering Research Council of Canada to P.T.; H.T.P. and O.T.H.H. acknowledge tuition fee waivers at INRS.

REFERENCES

1. Rivers CF, Longworth JF. 1972. A nonoccluded virus of *Junonia coenia* (*Nymphalidae*: *Lepidoptera*). J. Invertebr. Pathol. 20:369–370.
2. Longworth JF. 1978. Small isometric viruses of invertebrates. Adv. Virus Res. 23:103–157.
3. Jousset FX, Jourdan M, Compagnon B, Mialhe E, Veyrunes JC, Bergoin M. 1990. Restriction maps and sequence homologies of two densovirus genomes. J. Gen. Virol. 71(Pt 10):2463–2466.
4. Jourdan M, Jousset FX, 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 179:403–409.
5. Dumas B, Jourdan M, Pascaud AM, Bergoin M. 1992. Complete nucleotide sequence of the cloned infectious genome of *Junonia coenia* densovirus reveals an organization unique among parvoviruses. Virology 191: 202–222.
6. 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.
7. Fédière G, El-Far M, Li Y, Bergoin M, Tijssen P. 2004. Expression strategy of densonucleosis virus from *Mythimna loreyi*. Virology 320: 181–189.
8. 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. doi: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. Huang X, Madan A. 1999. CAP3: a DNA sequence assembly program. Genome Res. 9:868–877.
11. Wang Y, Abd-Alla AM, Bossin H, Li Y, Bergoin M. 2013. Analysis of the transcription strategy of the *Junonia coenia* densovirus (JcDNV) genome. Virus Res. 174:101–107.
12. Zádori Z, Szelei J, Lacoste MC, Li Y, Gariépy S, Raymond P, Allaire M, Nabi IR, Tijssen P. 2001. A viral phospholipase A_2 is required for parvovirus infectivity. Dev. Cell 1:291–302.

Annexe 6. Publication: *Acheta domesticus* volvovirus, a novel single-stranded circular DNA virus of the house cricket

genomeA
Journals.ASM.org

Acheta domesticus Volvovirus, a Novel Single-Stranded Circular DNA Virus of the House Cricket

Hanh T. Pham, Max Bergoin, Peter Tijssen

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

The genome of a novel virus of the house cricket consists of a 2,517-nucleotide (nt) circular single-stranded DNA (ssDNA) molecule with 4 open reading frames (ORFs). One ORF had a low identity to circovirus nucleotide sequences (NS). The unique properties of this volvovirus suggested that it belongs to a new virus family or genus.

Received 1 February 2013 **Accepted** 14 February 2013 **Published** 14 March 2013

Citation 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.

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

Cricket-breeding facilities in the United States produce billions of pet-feeder crickets annually (1, 2). The preferred house cricket, *Acheta domesticus*, is highly susceptible to a densovirus, *Acheta domesticus* densovirus (AdDNV), which has caused severe outbreaks since September 2009 and decimated *A. domesticus* stocks in North America. Samples received from die-offs were invariably positive for this virus. However, some recently received samples from mass cricket die-offs in North America were negative for AdDNV.

AdDNV-negative crickets (20 g) were homogenized in 20 ml of a 3:1 mixture of phosphate-buffered saline (PBS) and carbon tetrachloride. After low-speed centrifugation, the upper aqueous phase was passed through 0.45-nm filters and putative viruses were pelleted by centrifugation for 1.5 h at 40,000 rpm and resuspended in a small volume of Tris-EDTA (TE) buffer followed by DNase A and RNase A treatments to remove contaminating nucleic acids. Electron microscopy examination of a 100-fold dilution of the resuspended pellet revealed highly concentrated icosahedral particles of about 18 nm in diameter.

DNA extracted from purified virus by the High Pure viral nucleic acid kit (Roche Applied Science) was resistant to restriction endonucleases and presumably single stranded. Native viral DNA was used for double-stranded DNA synthesis at 30°C by ϕ29 DNA polymerase (3). Amplified DNA was digested with MboI, cloned into the BamHI site of the pBluescriptSK(–) vector, and sequenced by Sanger's method and primer walking as described before (4). The sequences were assembled by the CAP3 program (5) and generated a 2,517-nucleotide (nt) sequence containing a single EcoRI site. PCR using native DNA and 2 sets of outward primers (with respect to the EcoRI fragment) and sequencing confirmed the circular nature of the genome and the size of 2,517 nt. Due to the circular (rolling) nature of the genome, the name *Acheta domesticus* volvovirus (AdVVV; Volvo [Latin] = roll) was proposed.

Numbering of the genome started with the putative nonanucleotide origin of replication (1-TAGTATTAC), located, as for circo- or cycloviruses (6), between the open reading frames (ORFs) with opposite orientations. Among ORF products of >100 amino acids (aa), ORF1 (361 aa, starting at nt 447) and ORF4 (130 aa, starting at nt 70) were in the sense direction, whereas ORF2 (270 aa, starting at nt 2445) and the overlapping ORF3 (207 aa, starting at nt 2393) were in the antisense direction. BLASTn failed to detect any identity to viral sequences. However, BLASTp revealed a maximum identity of about 30% between ORF2 and Rep proteins of circoviruses and cycloviruses, with a coverage of ~85% (aa 5 to 80, Viral_Rep superfamily [pfam02407], and aa 150 to 212, P-loop_NTPase [pfam00910]). The other ORFs did not have any viral identity using BLASTp.

The lack of sequence identity and the differences in genome organization and size indicated a new virus family or genus. To our knowledge, this is the first circular single-stranded DNA virus in insects that is not related to cycloviruses (7, 8), circoviruses (9–11), nanoviruses (12, 13), or geminiviruses (14, 15), and it may be of interest in elucidating the evolution of this rapidly expanding virus group.

Nucleotide sequence accession number. The GenBank accession number for AdVVV is KC543331.

ACKNOWLEDGMENTS

This work was supported by funds from the Natural Sciences and Engineering Research Council of Canada to P.T.; H.T.P. acknowledges tuition waivers at INRS.

REFERENCES

1. 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.
2. 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.
3. 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. Proc. Natl. Acad. Sci. U. S. A. 107:16506–16511.
4. Tijssen P, Li Y, El-Far M, Szelei J, Letarte M, Zádori Z. 2003. Organi-

zation and expression strategy of the ambisense genome of densonucleosis virus of Galleria mellonella. J. Virol. 77:10357–10365.
5. Huang X, Madan A. 1999. CAP3: a DNA sequence assembly program. Genome Res. 9:868–877.
6. 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.
7. Rosario K, Dayaram A, Marinov M, Ware J, Kraberger S, Stainton D, Breitbart M, Varsani A. 2012. Diverse circular ssDNA viruses discovered in dragonflies (Odonata: Epiprocta). J. Gen. Virol. 93:2668–2681.
8. Rosario K, Marinov M, Stainton D, Kraberger S, Wiltshire EJ, Collings DA, Walters M, Martin DP, Breitbart M, Varsani A. 2011. Dragonfly cyclovirus, a novel single-stranded DNA virus discovered in dragonflies (Odonata: Anisoptera). J. Gen. Virol. 92:1302–1308.
9. 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.
10. Johne R, Fernández-de-Luco D, Höfle U, Müller H. 2006. Genome of a novel circovirus of starlings, amplified by multiply primed rolling-circle amplification. J. Gen. Virol. 87:1189–1195.
11. Kapoor A, Dubovi EJ, Henriquez-Rivera JA, Lipkin WI. 2012. Complete genome sequence of the first canine circovirus. J. Virol. 86:7018.
12. Timchenko T, Katul L, Aronson M, Vega-Arreguín JC, Ramirez BC, Vetten HJ, Gronenborn B. 2006. Infectivity of nanovirus DNAs: induction of disease by cloned genome components of Faba bean necrotic yellows virus. J. Gen. Virol. 87:1735–1743.
13. Timchenko T, Katul L, Sano Y, de Kouchkovsky F, Vetten HJ, Gronenborn B. 2000. The master rep concept in nanovirus replication: identification of missing genome components and potential for natural genetic reassortment. Virology 274:189–195.
14. Buchmann RC, Asad S, Wolf JN, Mohannath G, Bisaro DM. 2009. Geminivirus AL2 and L2 proteins suppress transcriptional gene silencing and cause genome-wide reductions in cytosine methylation. J. Virol. 83:5005–5013.
15. Martin DP, Lefeuvre P, Varsani A, Hoareau M, Semegni JY, Dijoux B, Vincent C, Reynaud B, Lett JM. 2011. Complex recombination patterns arising during geminivirus coinfections preserve and demarcate biologically important intra-genome interaction networks. PLoS Pathog. 7:e1002203. doi:10.1371/journal.ppat.1002203.

Annexe 7. Publication: New volvovirus isolates from *Acheta domesticus* (Japan) and *Gryllus assimilis* (United States)

genomeA
Journals.ASM.org

New Volvovirus Isolates from *Acheta domesticus* (Japan) and *Gryllus assimilis* (United States)

Hanh T. Pham,[a] Hajime Iwao,[b] Max Bergoin,[a] Peter Tijssen[a]

INRS-Institut Armand-Frappier, Laval, QC, Canada[a]; Niigata City Aquarium, Nishifunami, Niigata, Japan[b]

A novel circular single-stranded DNA (ssDNA) virus, volvovirus, from the house cricket has been described recently. Here, we report the isolation of volvoviruses from *Acheta domesticus* in Japan and *Gryllus assimilis* in the United States. These *Acheta domesticus* volvovirus (AdVVV) isolates have genomes of 2,517 and 2,516 nucleotides (nt) and 4 large open reading frames (ORFs).

Received 22 April 2013 **Accepted** 26 April 2013 **Published** 6 June 2013

Citation 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. doi:10.1128/genomeA.00328-13.

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

Billions of pet-feeder crickets are produced annually (1, 2). Since September 2009, stocks of the preferred house cricket, *Acheta domesticus*, were decimated due to severe fatal outbreaks caused by a densovirus, the *Acheta domesticus* densovirus (AdDNV) (2, 3). Similarly, in Japan house cricket stocks were decimated since early in the summer of 2009 despite strict import and export regulations. To avoid heavy losses, many cricket producers switched to the Jamaican field cricket, *Gryllus assimilis* (Fabricius) (1).

Some recent samples contained a new single-stranded, circular DNA virus (4), volvovirus or *Acheta domesticus* volvovirus (AdVVV)-IAF, that is not related to cycloviruses (5, 6), circoviruses (7, 8, 9), nanoviruses (10, 11), or geminiviruses (12, 13, 14).

Electron microscopy of diluted extracts from recently obtained dead house crickets from Japan and dead Jamaican field crickets from U.S. producers revealed highly concentrated icosahedral particles of about 18 nm diameter. Virus and nucleic acid were extracted as described previously (4). The nucleic acid was resistant to RNase and restriction enzymes, suggesting a single-stranded DNA (ssDNA) genome. Native viral DNA was used for double-strand DNA synthesis at 30°C by phi29 DNA polymerase (3, 15). Amplified DNA was digested with MboI or EcoRI, cloned into the BamHI or EcoRI sites of the pBluescriptSK(-) vector, and sequenced by Sanger's method and primer walking as described before (4, 16). The sequences were assembled by the CAP3 program (17) and generated a 2,517-nucleotide (nt) sequence for the Japanese isolate (AdVVV-Japan) and a 2,516-nt sequence for the *Gryllus assimilis* isolate (AdVVV-Ga). Both were closely related to AdVVV-IAF (4) and contained a single EcoRI site. PCR using native viral DNA and 2 sets of outward primers (with respect to the EcoRI fragment) and sequencing confirmed the circular nature of the genome. Nucleotide numbering was as for AdVVV-IAF (4).

Among open reading frames (ORFs) coding for >100 amino acids (aa), ORF1 (361 aa, starting at nt 447), putatively coding for the capsid protein (CP), and ORF4 (130 aa, starting at nt 70) were in the sense direction, whereas ORF2 (270 aa, starting at nt 2445 for AdVVV-Japan but at nt 2444 for AdVVV-Ga) and the overlapping ORF3 (207 aa, starting at nt 2393 for AdVVV-Japan but at 2392 for AdVVV-Ga) were in the antisense direction. BLASTp revealed a maximum identity of about 30% between ORF2 and rep proteins of circoviruses and cycloviruses, with coverage of ~85%. Compared to the original AdVVV-IAF isolate, AdVVV-Japan contained 22 substitutions of which none were in the overlapping ORF2 and ORF3, and only one was in the intergenic regions. Surprisingly, the bulk of substitutions were in the putative CP (18, of which 14 were nonsynonymous or 4% of protein sequence), whereas the putative ORF4 protein contained 3 nonsynonymous substitutions. Also striking was the high arginine content of the N terminus of the CP. The AdVVV-Ga isolate, compared to AdVVV-IAF, contained 4 nonsynonymous substitutions in the capsid protein also found in the Japanese isolate and one additional nonsynonymous substitution. It also had a deleted T at position 1620 in the intergenic region.

In conclusion, volvoviruses seem to be widespread and to infect several species of crickets.

Nucleotide sequence accession numbers. The GenBank accession number of AdVVV-Ga is KC794539 and that of AdVVV-Japan is KC794540.

ACKNOWLEDGMENTS

This work was supported by the Natural Sciences and Engineering Research Council of Canada to P.T.; H.T.P. acknowledges tuition waivers at INRS.

REFERENCES

1. 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.
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. 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.
4. 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.
5. Rosario K, Dayaram A, Marinov M, Ware J, Kraberger S, Stainton D, Breitbart M, Varsani A. 2012. Diverse circular ssDNA viruses discovered in dragonflies (Odonata: Epiprocta). J. Gen. Virol. 93:2668–2681.
6. Rosario K, Marinov M, Stainton D, Kraberger S, Wiltshire EJ, Collings DA, Walters M, Martin DP, Breitbart M, Varsani A. 2011. Dragonfly cyclovirus, a novel single-stranded DNA virus discovered in dragonflies (Odonata: Anisoptera). J. Gen. Virol. 92:1302–1308.
7. Johne R, Fernández-de-Luco D, Höfle U, Müller H. 2006. Genome of a novel circovirus of starlings, amplified by multiply primed rolling-circle amplification. J. Gen. Virol. 87:1189–1195.
8. Kapoor A, Dubovi EJ, Henriquez-Rivera JA, Lipkin WI. 2012. Complete genome sequence of the first canine circovirus. J. Virol. 86:7018.
9. 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.
10. Timchenko T, Katul L, Aronson M, Vega-Arreguín JC, Ramirez BC, Vetten HJ, Gronenborn B. 2006. Infectivity of nanovirus DNAs: induction of disease by cloned genome components of Faba bean necrotic yellows virus. J. Gen. Virol. 87:1735–1743.
11. Timchenko T, Katul L, Sano Y, de Kouchkovsky F, Vetten HJ, Gronenborn B. 2000. The master rep concept in nanovirus replication: identification of missing genome components and potential for natural genetic reassortment. Virology 274:189–195.
12. 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.
13. Martin DP, Lefeuvre P, Varsani A, Hoareau M, Semegni JY, Dijoux B, Vincent C, Reynaud B, Lett JM. 2011. Complex recombination patterns arising during geminivirus coinfections preserve and demarcate biologically important intra-genome interaction networks. PLoS Pathog. 7:e1002203.
14. Buchmann RC, Asad S, Wolf JN, Mohannath G, Bisaro DM. 2009. Geminivirus AL2 and L2 proteins suppress transcriptional gene silencing and cause genome-wide reductions in cytosine methylation. J. Virol. 83: 5005–5013.
15. de Vega M, Lázaro JM, Mencía M, Blanco L, Salas M. 2010. Improvement of phi29 DNA polymerase amplification performance by fusion of DNA binding motifs. Proc. Natl. Acad. Sci. U. S. A. 107:16506–16511.
16. 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.
17. Huang X, Madan A. 1999. CAP3: A DNA sequence assembly program. Genome Res. 9:868–877.

Annexe 8. Publication: A circo-like virus isolated from *Penaeus monodon* shrimps

genomeA
Journals.ASM.org

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 circoviruses 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.

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 ø29 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 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 (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 (NSERCC) 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 NSERCC.

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.

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, Blanco 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.

Annexe 9. Publication: A novel ambisense densovirus, *Acheta domesticus* mini ambidensovirus, from crickets

genomeA
Journals.ASM.org

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.

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 volvoviruses, 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 pBlueskript 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.

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.

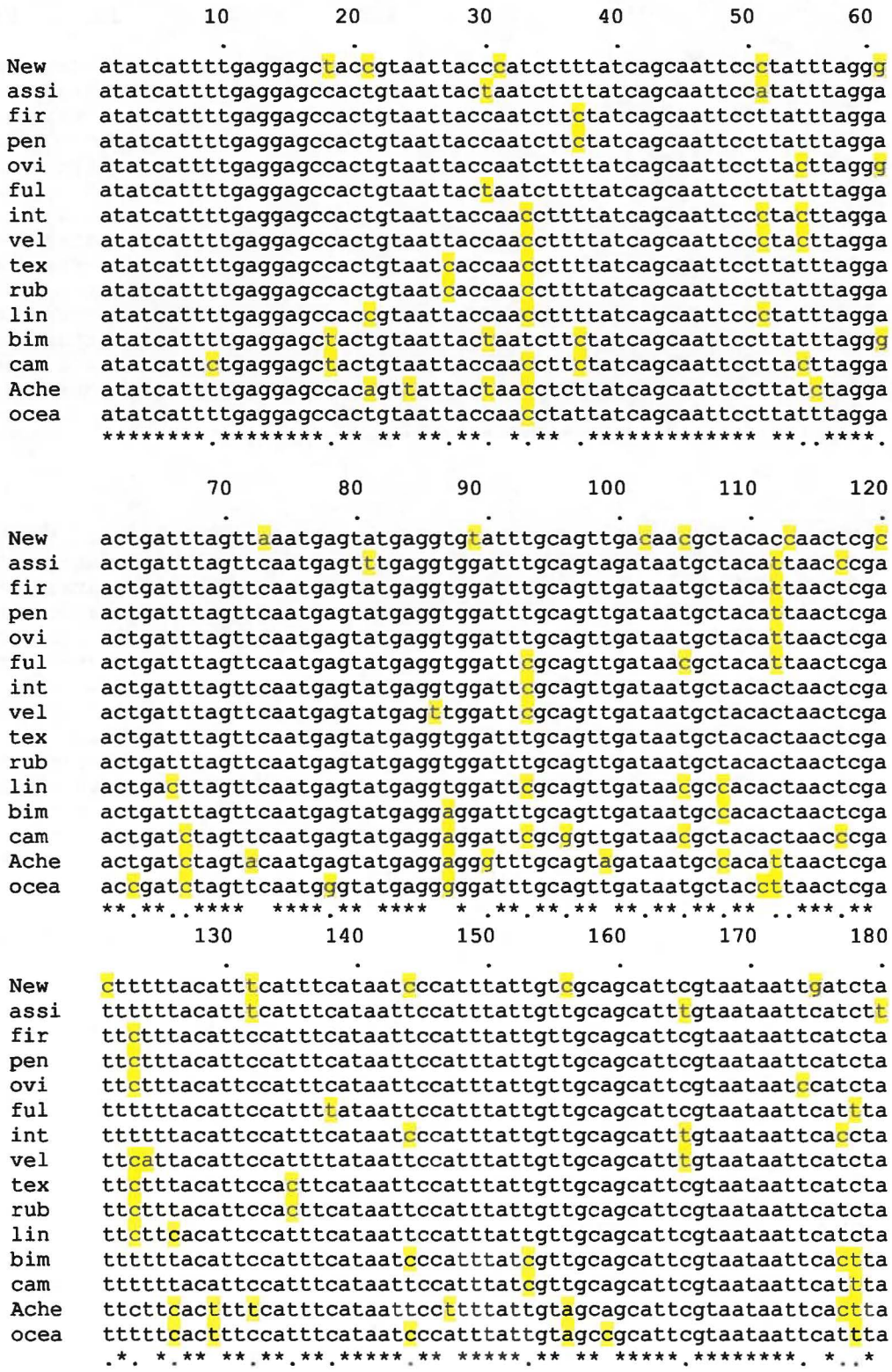
REFERENCES

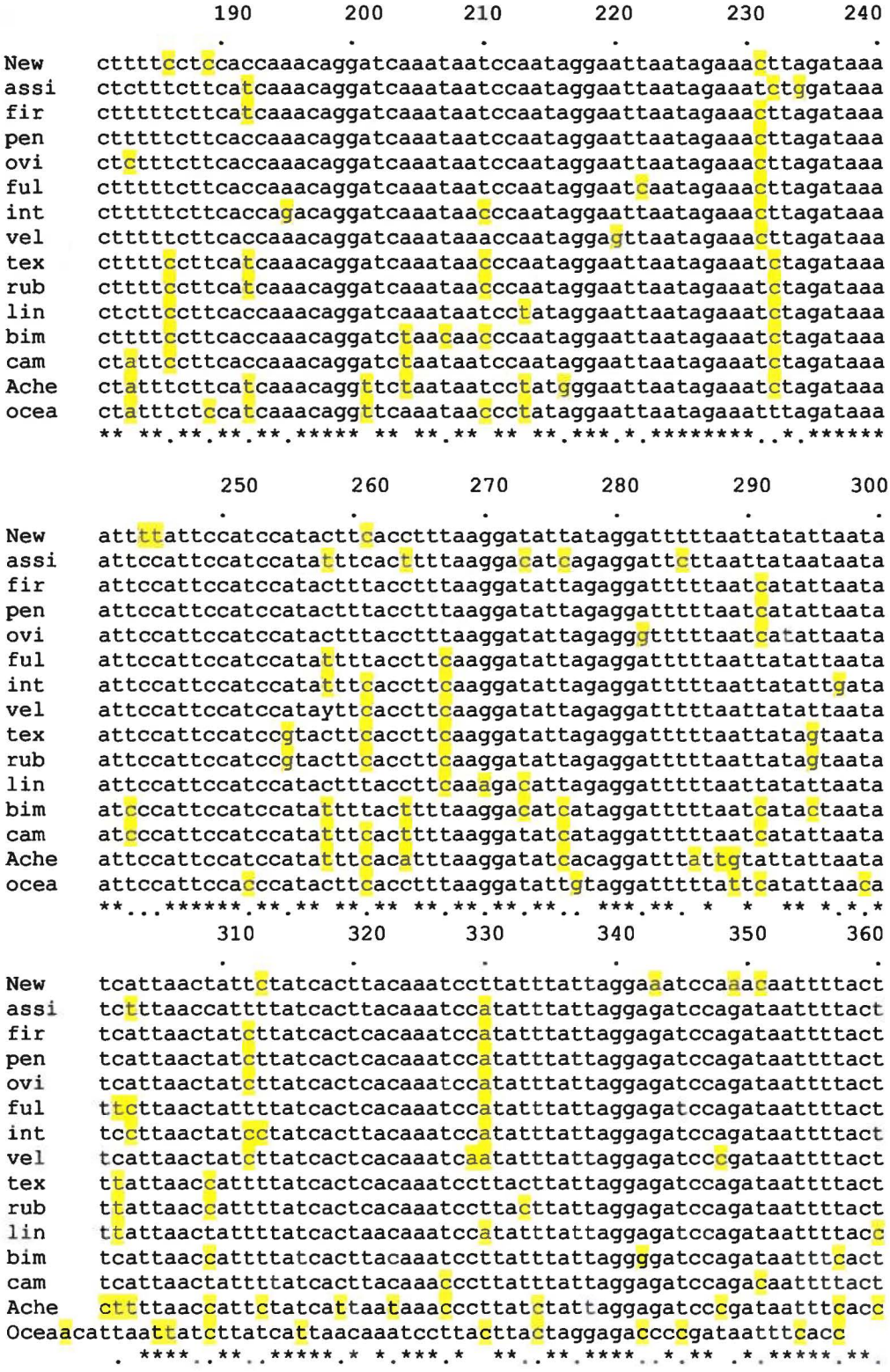
1. Liu K, Li Y, Jousset FX, Zadori Z, Szelei J, Yu Q, Pham HT, Lépine F, Bergoin M, Tijssen P. 2011. The *Acheta domesticus densovirus*, isolated from the European house cricket, has evolved an expression strategy unique among parvoviruses. J. Virol. 85:10069–10078.
2. Szelei J, Woodring J, Goettel MS, 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* densoviruses from different outbreaks in Europe, North America, and Japan. Genome Announc. 1(4):e00629-13. doi: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.
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.
8. **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. doi: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.

Annexe 10. Cricket species received for AdDNV diagnosis and comparison to other species

In 2009/2010, widespread epidemics of AdDNV wiped out a large part of the cricket industry with losses in the hundreds of millions of dollars (Szelei et al., 2010). Large cricket farms produce about 250 million crickets per year. Many farmers looked for alternatives and the crickets that we received, although often decomposed, seemed to be different from the house cricket. One particular type of sample contained what was called a "Crazy Red" or "New" sample. We decided to analyze the cytochrome b gene to see what was happening in the field. A 450 bp cytochrome b sequence of the "New", Ache (*Acheta domesticus*), assi (*Gryllus assimilis*), bim (*G. bimaculatus*) were determined and compared to fir (*G. firmus*), cam (*G. campestris*), int (*G. integer*), tex (*G. texensis*), ful (*G. fultoni*), lin (*G. lineaticeps*), ovi (*G. ovisopis*), rub (*G. rubens*), vel (*G. veletis*), pen (*G. pennsylvanicus*), ocea (*Teleogryllus oceanicus*) (Figure S1 and S2).


10 20 30 40 50 60
New atatcattttgaggagctaccgtaattacccatcttttatcagcaattccctatttaggg
assi atatcattttgaggagccactgtaattactaatcttttatcagcaattccatatttagga
fir atatcattttgaggagccactgtaattaccaatcttctatcagcaattccttatttagga
pen atatcattttgaggagccactgtaattaccaatcttctatcagcaattccttatttagga
ovi atatcattttgaggagccactgtaattaccaatcttttatcagcaattccttacttaggg
ful atatcattttgaggagccactgtaattactaatcttttatcagcaattccttatttagga
int atatcattttgaggagccactgtaattaccaaccttttatcagcaattccctacttagga
vel atatcattttgaggagccactgtaattaccaaccttttatcagcaattccctacttagga
tex atatcattttgaggagccactgtaatcaccaaccttttatcagcaattccttatttagga
rub atatcattttgaggagccactgtaatcaccaaccttttatcagcaattccttatttagga
lin atatcattttgaggagccaccgtaattaccaaccttttatcagcaattccctatttagga
bim atatcattttgaggagctactgtaattactaatcttctatcagcaattccttatttaggg
cam atatcattctgaggagctactgtaattaccaaccttctatcagcaattccttacttagga
Ache atatcattttgaggagccacagttattactaacctcttatcagcaattccttatctagga
ocea atatcattttgaggagccactgtaattaccaacctattatcagcaattccttatttagga
70 80 90 100 110 120
New actgatttagttaaatgagtatgaggtgtatttgcagttgacaacgctacaccaactcgc
assi actgatttagttcaatgagttgaggtggatttgcagtagataatgctacattaaccccga
fir actgatttagttcaatgagtatgaggtggatttgcagttgataatgctacattaactcga
pen actgatttagttcaatgagtatgaggtggatttgcagttgataatgctacattaactcga
ovi actgatttagttcaatgagtatgaggtggatttgcagttgataatgctacattaactcga
ful actgatttagttcaatgagtatgaggtggattcgcagttgataacgctacattaactcga
int actgatttagttcaatgagtatgaggtggattcgcagttgataatgctacactaactcga
vel actgatttagttcaatgagtatgagttggattcgcagttgataatgctacactaactcga
tex actgatttagttcaatgagtatgaggtggatttgcagttgataatgctacactaactcga
rub actgatttagttcaatgagtatgaggtggatttgcagttgataatgctacactaactcga
lin actgacttagttcaatgagtatgaggtggattcgcagttgataacgccacactaactcga
bim actgatttagttcaatgagtatgaggaggatttgcagttgataatgccacactaactcga
cam actgatctagttcaatgagtatgaggaggattcgcggttgataacgctacactaaccgga
Ache actgatctagtacaatgagtatgaggagggtttgcagtagataatgccacattaactcga
ocea accgatctagttcaatgggtatgaggggatttgcagttgataatgctaccttaactcga
130 140 150 160 170 180
New cttttacatttcatttcataatcccatttattgtcgcagcattcgtaataattgatcta
assi ttttttacatttcatttcataattccatttattgttgcagcatttgtaataattcatctt
fir ttctttacattccatttcataattccatttattgttgcagcattcgtaataattcatcta
pen ttctttacattccatttcataattccatttattgttgcagcattcgtaataattcatcta
ovi ttctttacattccatttcataattccatttattgttgcagcattcgtaataatccatcta
ful ttttttacattccatttttataattccatttattgttgcagcattcgtaataattcattta
int ttttttacattccatttcataatcccatttattgttgcagcatttgtaataattcacctа
vel ttcattacattccatttttataattccatttattgttgcagcatttgtaataattcatcta
tex ttctttacattccacttcataattccatttattgttgcagcattcgtaataattcatcta
rub ttctttacattccacttcataattccatttattgttgcagcattcgtaataattcatcta
lin ttcttcacattccatttcataattccatttattgttgcagcattcgtaataattcatcta
bim ttttttacattccatttcataatcccatttatcgttgcagcattcgtaataattcactta
cam ttttttacattccatttcataattccatttatcgttgcagcattcgtaataattcattta
Ache ttcttcacttttcatttcataattccttttattgtagcagcattcgtaataattcactta
ocea tttttcactttccatttcataatcccatttattgtagccgcattcgtaataattcattta


190 200 210 220 230 240
New ctttt cctc caccaaacaggatcaaataatccaataggaattaatagaaac ttagataaa
assi ctctttcttcatcaaacaggatcaaataatccaataggaattaatagaaatctggataaa
fir ctttttcttcatcaaacaggatcaaataatccaataggaattaatagaaac ttagataaa
pen ctttttcttcaccaaacaggatcaaataatccaataggaattaatagaaac ttagataaa
ovi ctctttcttcaccaaacaggatcaaataatccaataggaattaatagaaac ttagataaa
ful ctttttcttcaccaaacaggatcaaataatccaataggaatcaatagaaac ttagataaa
int ctttttcttcaccagacaggatcaaataacccaataggaattaatagaaac ttagataaa
vel ctttttcttcaccaaacaggatcaaataaaccaataggagttaatagaaac ttagataaa
tex ctttt ccttcatcaaacaggatcaaataacccaataggaattaatagaaatctagataaa
rub ctttt ccttcatcaaacaggatcaaataacccaataggaattaatagaaatctagataaa
lin ctcttccttcaccaaacaggatcaaataatcctataggaattaatagaaatctagataaa
bim ctttt ccttcaccaaacaggatctaacaacccaataggaattaatagaaatctagataaa
cam ctattccttcaccaaacaggatctaataatccaataggaattaatagaaatctagataaa
Ache ctatttcttcatcaaacaggttctaataatcctatgggaattaatagaaatctagataaa
ocea ctatttctccatcaaacaggttcaaataaccctataggaattaatagaaatttagataaa
** **.**.**.**.***** ** **.** ** **.***.*.********..*.******
250 260 270 280 290 300
New atttt attccatccatacttcacctttaaggatattataggatttttaattatattaata
assi attccattccatccatatttcactttaaggacatcagaggattcttaattataataata
fir attccattccatccatactttacctttaaggatattagaggatttttaatcatattaata
pen attccattccatccatactttacctttaaggatattagaggatttttaatcatattaata
ovi attccattccatccatactttacctttaaggatattagagggttttttaatcatattaata
ful attccattccatccatatttttaccttcaaggatattagaggatttttaattatattaata
int attccattccatccatatttcaccttcaaggatattagaggatttttaattatattgata
vel attccattccatccataytt caccttcaaggatattagaggatttttaattatattaata
tex attccattccatccgtacttcaccttcaaggatattagaggatttttaattatagtaata
rub attccattccatccgtacttcaccttcaaggatattagaggatttttaattatagtaata
lin attccattccatccatactttaccttcaaagacattagaggatttttaattatattaata
bim atcccattccatccatattttacttttaaggacatcataggatttttaatcatactaata
cam atcccattccatccatatttcacttttaaggatatcataggatttttaatcatattaata
Ache attccattccatccatatttcacatttaaggatatcacaggatttattgtattattaata
ocea attccattccacccatacttcacctttaaggatattgtaggatttttattcatattaaca
...****.**.** **.** **.**.**.**.. ***.**. * * ** *.*.*
310 320 330 340 350 360
New tcattaactattctatcacttacaaatccttatttattaggaaatccaaacaatttttact
assi tctttaaccatttttatcacttacaaatccatatttattaggagatccagataatttttact
fir tcattaactatcttatcactcacaaatccatatttattaggagatccagataatttttact
pen tcattaactatcttatcactcacaaatccatatttattaggagatccagataatttttact
ovi tcattaactatcttatcactcacaaatccatatttattaggagatccagataatttttact
ful ttcttaactattttatcactcacaaatccatatttattaggagatccagataatttttact
int tccttaactatcctatcacttacaaatccatatttattaggagatccagataatttttact
vel tcattaactatcttatcactcacaaatcaatatttattaggagatcccgataatttttact
tex ttattaaccattttatcactcacaaatccttacttattaggagatccagataatttttact
rub ttattaaccattttatcactcacaaatccttacttattaggagatccagataatttttact
lin ttattaactattttatcactaacaaatccatatttattaggagatccagataatttttacc
bim tcattaaccattttatcacttacaaatccttatttattaggggatccagataatttcact
cam tcattaactattttatcacttacaaaccctttatttattaggagatccagacaatttttact
Ache ctttt aaccattctatcattaataaaccctttatctattaggagatcccgataatttcacc
Oceaacattaattatcttatcattaacaaatccttacttactaggagaccccgataatttcacc
. ****..**..*****.* *.***.* **..**.****..*.** .*.*****.**.

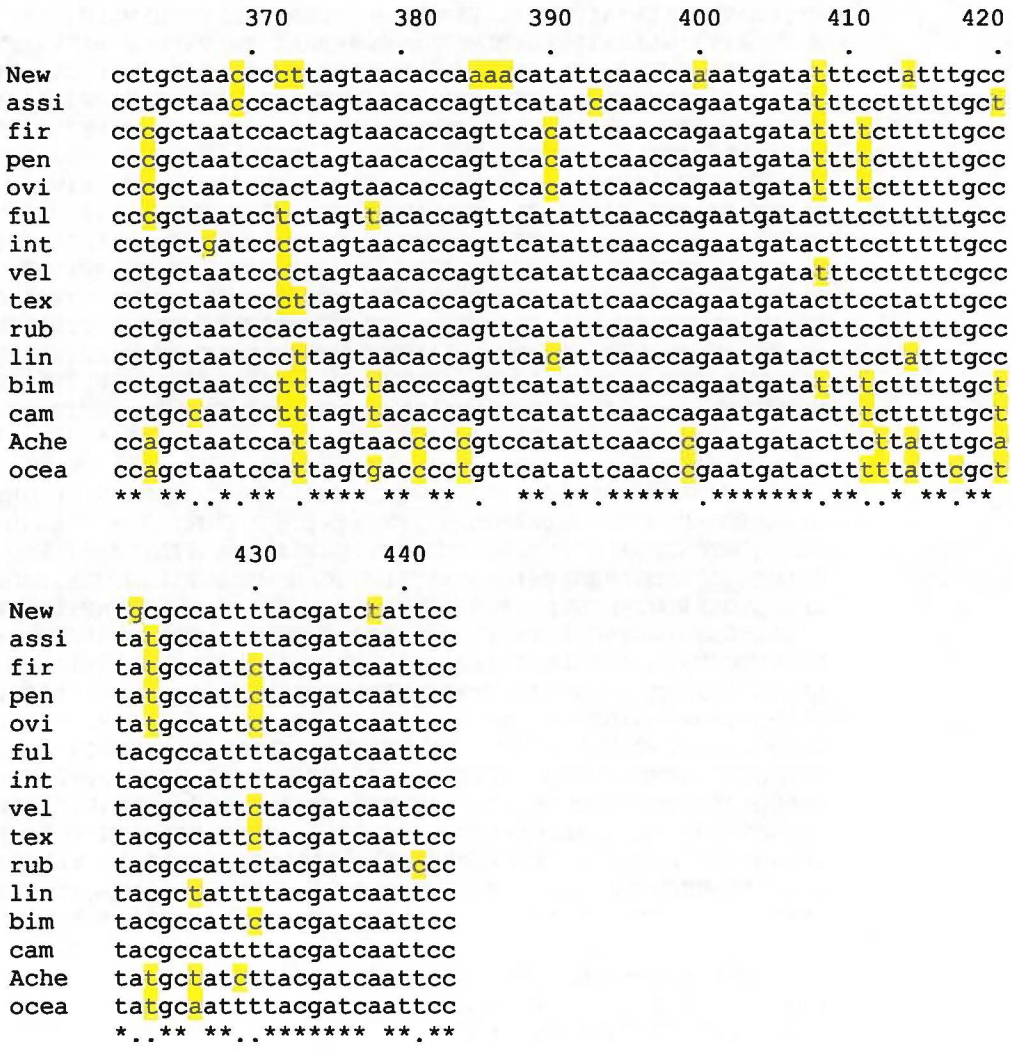


Figure S1. Analysis of DNA sequence polymorphisms of 450-bp cytochrome b gene. Polymorphisms are highlighted. Clustal format alignment was with MAFFT (v7.130b), "1PAM/K=2" scoring matrix for nucleotide sequences.

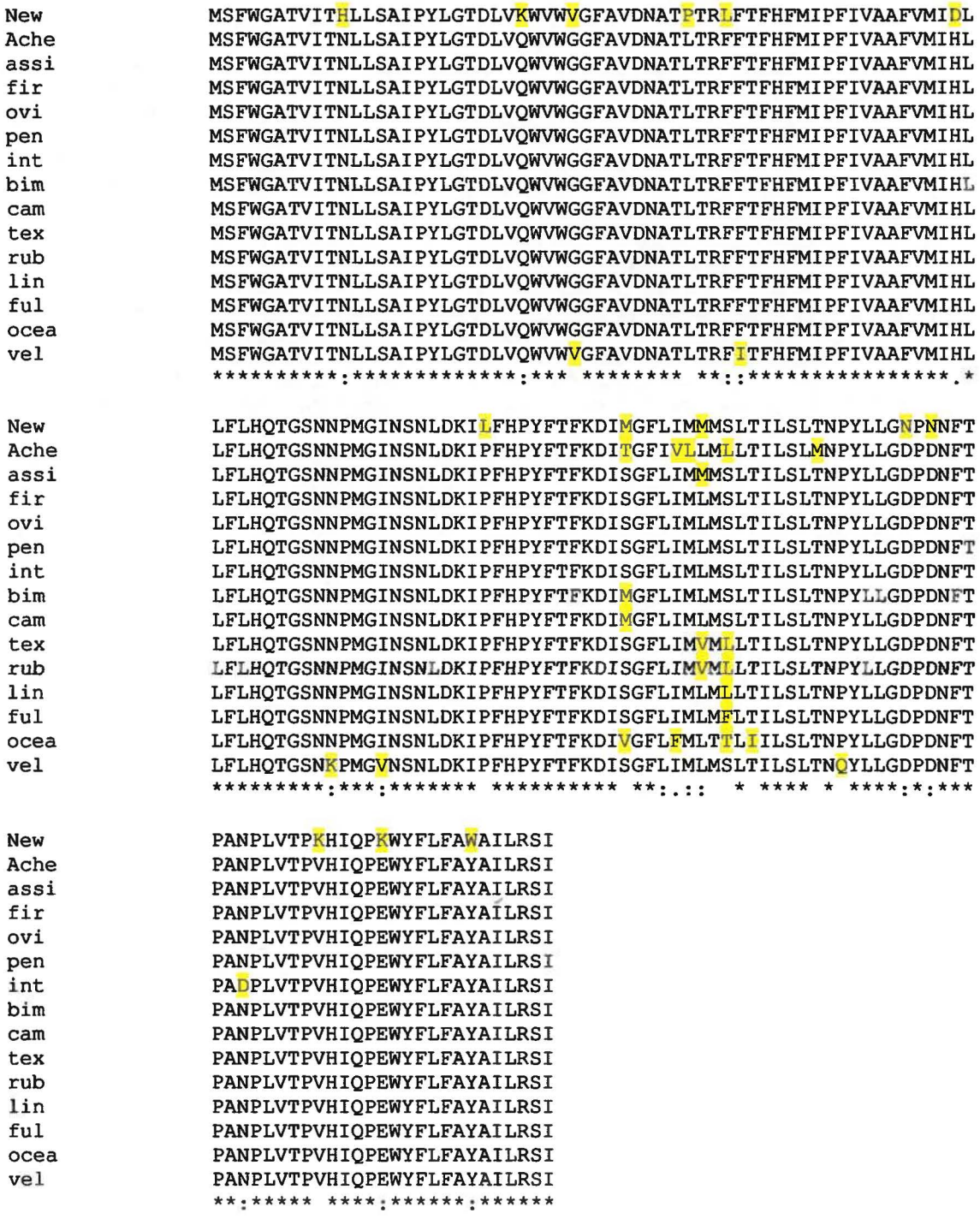


Figure S2. Protein alignment of CLUSTAL format alignment by MAFFT (v7.130b) Scoring matrix for amino acid sequences BLOSUM62.

Publication: La vente de milliards de grillons (Orthoptères de la famille des Gryllidae) servant au nourrissage des animaux de compagnie, les fermes industrielles à grillons, un densovirus épizootique et les réglementations gouvernementales contribuent à un désastre potentiel. David B. Weissman, David A. Gray, Hanh Thi Pham & Peter Tijssen, Zootaxa 2012;3504: 67-88.

L'élevage industriel de grillons servant au nourrissage d'animaux de compagnie ou comme appâts produit environ 50 millions de grillons par semaine aux États-Unis. Les fermes à grillons et leur commercialisation emploient une main d'oeuvre importante et représente un chiffre d'affaires de plusieurs centaines millions de dollars par an. Ce commerce s'est effondré dans les années 2009-2010 par suite d'épizooties dévastatrices dans tout le continent nord-américain. En dépit des lois américaines sur l'importation d'insectes, plusieurs espèces de grillons supposées résistantes au virus ont été introduites, certaines illégalement, et commercialisées. Cette Publication présente les résultats de recherches faites en collaboration entre le Département d'entomologie de l'Académie des Sciences de Californie à San Francisco, le Département de Biologie de l'Université d'État de Californie à Northridge, notre laboratoire et les principaux producteurs. Ces recherches ont eu pour but d'identifier les espèces de grillons introduites: *Gryllodes sigillatus*, *Gryllus assimilis*, *Gryllus bimaculatus* et une nouvelle espèce, *Gryllus locorojo*, de faire une analyse épidémiologique de l'AdDNV dans plusieurs centaines d'échantillons de ces différentes espèces et d'étudier leur sensibilité à ce virus. L'ensemble de ces travaux a permis de formuler des recommandations auprès des producteurs et de sensibiliser les services gouvernementaux sur les risques potentiels de ces introductions. Nos travaux ont montré que la plupart des échantillons d'*A. domesticus* étaient infectés par l'AdDNV et que les espèces introduites sont également sensibles au virus mais à un degré moindre. Des analyses en PCR semi-quantitative ont confirmé que l'AdDNV se réplique au moins chez les deux espèces *G. assimilis* et *G. locorojo*. Par contre la présence d'AdDNV dans *G. bimaculatus* n'est détecTable qu'aux plus faibles dilutions des spécimens.

Contribution to this Publication: A collaboration between the Department of Entomology, California Academy of Sciences (San Francisco), Department of Biology, California State University (Northridge) and our laboratory. Pham's contribution was the virology work and initial work on mitochondrial DNA analysis of different cricket species. Thereafter, the other labs were involved because of their expertise in cricket taxonomy. Pham wrote the virology part of the manuscript. Weissman and Tijssen supervised the project.

Publication: Billions and billions sold: Pet-feeder crickets (*Orthoptera*: *Gryllidae*), commercial cricket farms, an epizootic densovirus, and government regulations make for a potential disaster

Zootaxa 2012 Oct; 3504: 67–88

David B. Weissman[1], David A. Gray[2], Hanh Thi Pham[3]& Peter Tijssen[3]

[1] Department of Entomology, California Academy of Sciences, San Francisco, CA94118

[2] Department of Biology, California State University, Northridge, CA 91330

[3] INRS-Institut Armand-Frappier, Laval, QC, Canada H7V 1B7