

Université du Québec
INRS-Institut Armand-Frappier

***Pseudoplusia includens* densovirus (PiDNV)
genome organization and expression strategy**

Par
Oanh Thi Hoang Huynh

Mémoire présenté pour l'obtention du grade de Maître ès sciences (M.Sc.)
en Virologie et Immunologie

Jury d'évaluation

Examineur interne	Angela Pearson, Ph.D. INRS-Institut Armand-Frappier
Examineur externe	Guy Lemay, Ph.D. Département de microbiologie et immunologie Université de Montréal
Directeur de recherche	Peter Tijssen, Ph.D. INRS-Institut Armand-Frappier
Co-directeur de recherche	Tsuneyuki Ozaki, Ph.D. INRS-Énergie, Matériaux et Télécommunication

ACKNOWLEDGEMENTS

Last two years and a half is one of the most special journeys that I have ever experienced. I would not have been possible to go to the end without the support of many kind people to whom I would like to show my deepest gratitude.

I am truly thankful to the financial support of the Biotechnology center of Ho Chi Minh city and its scholarship program, which gave me the opportunity to do my Master in Canada. I would also like to thank the financial and academic support of the Armand-Frappier Institute, where I spent most of my studying time, and Dr. Ozaki (INRS-EMT) for financially supporting my virus inactivation work.

I owe sincere and earnest thankfulness to my supervisor, Peter Tijssen and my co-supervisor, Tsuneyuki Ozaki who kindly gave me the chance to work in their laboratories and guided me throughout my study. I am sure that I would not have been able to finish without their support, inspiration, understanding and encouragement.

I am also sincerely grateful to the scientific support of my colleagues who I feel very lucky to work with. Your support really helped me develop my techniques as well as acquire knowledge in the laboratory.

I am obliged to my family and my friends whose constantly moral support has been making my life more enjoyable.

It is again a great pleasure to acknowledge everyone who has ever helped me throughout my study.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS	<i>i</i>
TABLE OF CONTENTS	<i>ii</i>
LIST OF FIGURES	<i>v</i>
LIST OF TABLES	<i>vii</i>
LIST OF ABBREVIATIONS	<i>viii</i>
SOMMAIRE RÉCAPITULATIF	1
INTRODUCTION	6
REVIEW OF LITERATURE	8
1. The family <i>Parvoviridae</i>	9
1.1. Classification.....	9
1.2. Virus particle	11
1.2.1. Viral capsid.....	11
1.2.2. Viral genome	12
1.2.2.1. Telomeres	13
1.2.2.2. Genome organization of subgroup A ambisense densoviruses	14
1.3. Viral life cycle.....	16
2. <i>Pseudoplusia includens</i> densovirus (PiDNV)	18
METHODOLOGY	20
1. Cloning and sequencing the genome of PiDNV	21
1.1. Isolation and purification of PiDNV DNA from infected larvae.....	21
1.2. Digestion of PiDNV DNA with the ClaI enzyme.....	21

1.3. Agarose gel electrophoresis	21
1.4. Preparation of insert	22
1.5. Preparation of vector	22
1.6. Ligation	23
1.7. Transformation	24
1.7.1. Preparation of electrocompetent cells	24
1.7.2. Electroporation	24
1.8. Plasmid DNA extraction	25
1.9. Analysis of clones by restriction digestion	25
1.10. Sequencing	26
1.11. Cloning of the whole genome of PiDNV	27
1.12. Sequencing of hairpins in linear pJAZZ clones	28
2. Bioinformatics	29
2.1. Primer design	29
2.2. Assembly of sequence of PiDNV genome.....	29
2.3. Structure analysis of viral hairpins.....	29
2.4. Genome sequence analysis.....	29
2.5. Prediction of 3D structure of PiDNV capsid subunit and capsid.....	30
RESULTS	31
1. Cloning of PiDNV genome	32
1.1. Isolation and purification of PiDNV DNA	32
1.2. Cloning of two halves of PiDNV genome	32
1.3. Cloning of viral hairpins	33
1.3.1. Subcloning BamHI-BamHI fragment into pBluescript KS (+).....	33

1.3.2. Cloning of two parts of hairpins.....	35
1.4. Cloning the whole genome of PiDENV	36
1.5. PCR to amplify the 5'-hairpin.....	38
2. Sequence and organization of PiDENV genome.....	39
2.1. Telomeres of PiDENV genome.....	39
2.2. Coding sequences of PiDENV genome.....	42
2.2.1. Non-structural proteins.....	43
2.2.2. Structural proteins	44
3. Predicted structure of capsid subunit and capsid of PiDENV	46
DISCUSSION	51
1. Cloning and sequencing of PiDENV genome.....	52
2. Nucleotide sequence of PiDENV genome and comparison with other viruses in the family.....	53
3. Genome organization of PiDENV and comparison with other densoviruses	53
4. Prediction of the 3D structure of the capsid subunit and capsid of PiDENV	55
CONCLUSION	56
APPENDIX 1. PUBLICATION.....	58
APPENDIX 2. VIRUS INACTIVATION BY IMPULSIVE STIMULATED RAMAN SCATTERING (ISRS).....	62
REFERENCES.....	87

LIST OF FIGURES

Figure 1. Phylogenetic relationship among the pleiotropic NS1 proteins of members of the family <i>Parvoviridae</i>	10
Figure 2. Morphology of parvoviruses	11
Figure 3. Genome organization of parvoviruses	12
Figure 4. Genome organization of ambisense densoviruses	15
Figure 5. Terminal hairpin of subgroup A densoviruses	15
Figure 6. Parvoviral "rolling-hairpin" DNA replication.....	18
Figure 7. DNA of PiDNV	19
Figure 8. Vector pBluescript KS (+)	23
Figure 9. The pJAZZ-OC vector	28
Figure 10. Purified PiDNV DNA	32
Figure 11. Clones of two halves of PiDNV genome (digested with XbaI).....	33
Figure 12. Cloning for sequencing of hairpins.....	34
Figure 13. BamHI-BamHI subclone.....	34
Figure 14. Digestion of BamHI-BamHI clones.....	35
Figure 15. Hairpin clones (digested with PvuII)	36
Figure 16. Example of colony PCR of the cloning of the whole genome of PiDNV	37
Figure 17. Positive linear clone of complete PiDNV genome (NotI)	37
Figure 18. First PCR of the 5'-hairpin from linear clones	38
Figure 19. Second (nested) PCR of the 5'-hairpin from linear clones.....	38
Figure 20. Alignment of two versions (flip and flop) of 3'-terminal hairpin.....	40
Figure 21. Hairpin of PiDNV	41
Figure 22. Genome organization	43

Figure 23. Conserved domain within the ORF2.....	45
Figure 24. Alignment of the conserved domain within ORF4 (30483) with the conserved domain of capsid protein VP4 of densoviruses.....	46
Figure 25. Alignment of phospholipase A ₂ domain in VP1 of PiDNV and representative parvoviruses	46
Figure 26. Alignment of amino acid sequence of VP of MVM and PPV	47
Figure 27. Alignment of amino acid sequence of VP4 of PiDNV and GmDNV (COBALT)....	48
Figure 28. Example of co-ordinates of last atoms in the protein string of capsid subunit of PiDNV	48
Figure 29. Reliability of PiDNV capsid subunit model: QMEAN4 global scores.....	48
Figure 30. Reliability of PiDNV capsid subunit model according to Swiss-Model (online): Colouring by residue error	49
Figure 31. PiDNV capsid subunit.....	49
Figure 32. Model of capsid structure of PiDNV	50
Figure 33. Viruses in inactivation experiments.....	70
Figure 34. Experimental setup.....	71
Figure 35. Example of a titration plate.....	75
Figure 36. Impact of time on virus inactivation	77
Figure 37. Impact of energy on virus inactivation	78
Figure 38. Impact of laser pulse width on virus inactivation	79
Figure 39. Comparison of remaining virus infectivity between one- and two-colour irradiation.	80
Figure 40. Virus inactivation with both 800 nm and 400 nm laser	81

LIST OF TABLES

Table 1.	Sequence of primers for sequencing PiDNV genome	26
Table 2.	BLAST Alignment of PiDNV sequence and representative parvovirus sequences ..	39
Table 3.	Predicted promoter sequence and transcription factors for these promoters	42
Table 4.	Selected protein motifs and profiles present in ORF1, ORF2, ORF3 and ORF4	45
Table 5.	Representative inactivation technologies	69
Table 6.	Experiments with the 10 Hz laser	74

LIST OF ABBREVIATIONS

aa	: Amino acid
AAV	: Adeno-associated virus
ATP	: Adenosine triphosphate
CDC	: Centers for Disease control and Prevention
CDD	: Conserved domains database
DMSO	: Dimethyl sulfoxide
DNA	: Deoxyribonucleic acid
dNTP	: Deoxyribonucleotide triphosphate
EDTA	: Ethylenediaminetetraacetic acid
fs	: Femtosecond
GmDNV	: <i>Galleria mellonella</i> densovirus
HBV	: Hepatitis B virus
HCV	: Hepatitis C virus
HIV	: Human immunodeficiency virus
HTLV	: Human T-lymphotropic virus
ITR	: Inverted terminal repeat
ISRS	: Impulsive Stimulated Raman Scattering
JcDNV	: <i>Junonia coenia</i> densovirus
LB	: Luria broth
MVM	: Minute virus of mice
NS	: Non-structural (protein)
NID	: Non-ionic detergent
ORF	: Open reading frame

- PCR** : Polymerase chain reaction
- Pfu** : Plaque forming units
- PfDNV** : Pefudensovirus
- PiDNV** : *Pseudoplusia includens* densovirus
- PLA₂** : Phospholipase A₂
- PPV** : Porcine parvovirus
- u** : Unit (of enzyme activity)
- VP** : Structural protein
- WNV** : West Nile virus
- 3D** : Three-dimensional

SOMMAIRE RÉCAPITULATIF

La famille de *Parvoviridae* regroupe des virus de petite taille (215-255 Å) sans enveloppe dont la capsid est de symétrie icosaédrique et dont le génome est une molécule d'ADN simple brin d'une taille comprise entre 4 et 6 kb. Cette famille se subdivise en deux sous-familles, celles des *Parvovirinae* regroupant les virus qui infectent les vertébrés et celle des *Densovirinae* regroupant les virus pathogènes d'invertébrés (68). Environ la moitié seulement des densovirus isolés à ce jour sont classés, faute de caractérisation suffisante de leur génome. Le densovirus de *Pseudoplusia includens* (PiDNV), objet de notre étude, a été isolé par Chao *et al.* (13) à partir des larves de ce lépidoptère. Les particules de PiDNV sont icosaédriques et contiennent une molécule d'ADN linéaire simple brin de 6 kb. À chaque extrémité de la molécule une répétition terminale inversée (ITR) correspondant à 6-7% de la taille du génome permet, par appariement des séquences complémentaires, la formation d'une structure en forme de « poêle à frire » (panhandle-like structure) observable au microscope électronique (13). L'électrophorèse en gel dénaturant SDS-polyacrylamide a révélé 4 protéines virales VP1-4 de 87 à 46 kDa. Par immunodiffusion, une réaction croisée partielle a été observée entre le virus GmDNV et le PiDNV (13). *Pseudoplusia includens*, l'hôte du PiDNV, a une aire de répartition s'étendant sur tout le continent américain (Nord et Sud). Il s'agit d'une espèce polyphage pouvant se nourrir non seulement sur le soja mais aussi d'autres plantes économiquement importantes comme la patate douce, l'arachide, le coton, la tomate, les crucifères, etc. (12). En conséquence, le virus PiDNV peut être utilisé comme un agent de lutte biologique contre ce ravageur. Cependant, depuis la première caractérisation de Chao *et al.* (13), aucune caractérisation moléculaire de son génome n'a été pas faite. Nous reportons dans ce mémoire les résultats du clonage et du séquençage du génome du PiDNV ainsi que l'analyse de son organisation et de sa stratégie d'expression.

Nous avons purifié le PiDNV à partir des larves de *Pseudoplusia includens*, qui ont été fournies par le Dr. Yu-Chan Chao en 2000. Le génome du virus a été extrait et son analyse en gel d'agarose a permis de visualiser d'une bande de 6 kb. Ce génome a été coupé en deux moitiés par l'enzyme de restriction ClaI et les deux moitiés ont été clonées dans le vecteur pBluescript KS (+) et séquencées. Le séquençage de la région interne du génome, entre les répétitions terminales inversées (ITRs), n'a pas posé de problème. Par contre, le séquençage des ITRs dont l'appariement de séquences complémentaires génère des structures en « épingles à cheveux » (hairpins) s'est avéré plus difficile et on a dû faire appel à différentes stratégies pour obtenir ces séquences. Ainsi, l'épingle à cheveux à chaque extrémité du génome viral dans un clone a été coupée au site de restriction BstUI en deux moitiés et chaque moitié a été clonée puis séquencée séparément. Cette approche a permis d'obtenir la séquence de l'extrémité 3' mais pas celle de l'extrémité 5'. Le génome complet de PiDNV a été également cloné dans le vecteur linéaire pJAZZ-OC spécialement adapté au clonage d'ADN difficile à cloner (31). Seulement 6 clones positifs ont été obtenus à partir de 1848 colonies (0.3%). Ces clones ont été séquencés et leur séquençage a confirmé la séquence obtenue avec les séquences clonées dans pBluescript sans toutefois fournir la séquence de l'épingle à cheveux à l'extrémité 5'. Nous avons alors amplifié par PCR les ITRs en présence de trois additifs: la bétaine, le DMSO et le 7-deaza dGTP

(47). Cette méthode a permis d'obtenir la séquence de l'ITR 5'. L'assemblage à l'aide du programme CAP3 de tous les contigs a généré une séquence de 5990 nucléotides correspondant au génome complet du PiDnV. BLAST a démontré une identité élevée entre le génome de PiDnV et les génomes des virus du sous-groupe A du genre *Densovirus*: JcDnV (87%), MIDnV (86%) and GmDnV (84%) et une identité beaucoup plus faible (<25%) avec les séquences des autres densovirus, y compris ceux appartenant aux autres sous-groupes du même genre. L'ITR 5' terminal et l'ITR 3' terminal du PiDnV ont chacun une taille de 540 nucléotides et des séquences parfaitement complémentaires l'une de l'autre. Les 120 nucléotides aux extrémités de ces ITRs peuvent se replier pour créer les épingles à cheveux ayant une configuration «Y», semblable à celles des virus MIDnV, GmDnV et JcDnV (66). Comme chez ces derniers virus, les configurations «flip» et «flop» sont aussi observées dans les épingles à cheveux de PiDnV. Dans ces deux configurations, les séquences formant la «tige» (stem) de l'«Y» sont identiques tandis que les deux petites palindromiques formant les deux «bras» de l'«Y» sont inversées dans la configuration «flip» par rapport à la configuration «flop». On a retrouvé également près de l'extrémité de 3' des épingles à cheveux le motif hautement conservé (GAC)₄ (24) chez tous les membres du sous-groupe A du genre *Densovirus* qui est considéré comme le site de fixation de la protéine NS1 pour couper le génome et produire l'origine de réplication. On a trouvé également dans la séquence des ITRs, les séquences spécifiques des deux promoteurs régulant l'expression des gènes non-structuraux (NS) sur un brin et structuraux (VP) sur le brin complémentaire avec leur «TATA box» localisées juste à la fin des ITRs.

Le programme ORF Finder (NCBI) a identifié quatre cadres ouverts de lecture (ORFs) dans le génome de PiDnV (ORF1-4). ORF1, ORF2 et ORF3 sont localisés sur la moitié 5' d'un brin tandis que l'ORF4 occupe la moitié 5' du brin complémentaire. Cette organisation est très semblable à celle du virus GmDnV, un densovirus ambisense dans le sous-groupe A du genre *Densovirus*. La distribution des ORF2, ORF3 et ORF1 de PiDnV correspond respectivement à la distribution des protéines non-structurales NS1, NS2 et NS3 du virus GmDnV. Il en est de même, pour l'ORF4 du PiDnV dont l'organisation correspond à l'ORF des protéines VP du virus GmDnV.

Les séquences codantes des protéines NS de PiDnV sont portées par les ORF1, ORF2 et ORF3. Le premier ORF (ORF1, nt 647 à 1348) code pour une protéine de 233 aa dont la séquence montre une forte homologie avec la séquence de la protéine NS3 des virus MIDnV (75%), GmDnV (73%) et JcDnV (72%). Le second ORF (ORF2, nt 1355 à 3019) code pour une protéine de 554 aa. Cette protéine possède un domaine hautement conservé de la protéine non-structurale NS1 des parvovirus. Le troisième ORF (ORF3, nt 1362 à 2189) est inclus dans la séquence de l'ORF2 et code pour une protéine de 275 aa présentant une très forte homologie avec la protéine non-structurale NS3 des virus MIDnV (96%), JcDnV (93%) et GmDnV (94%). La forte homologie de séquence des génomes du PiDnV et du GmDnV et l'organisation similaire de leurs séquences codantes suggèrent que PiDnV utilise la même stratégie

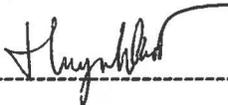
d'expression que GmDENV. Ainsi ORF1 du PiDENV qui porte la séquence NS3 est selon toute vraisemblance traduit à partir d'un transcrit NS épissé de la séquence NS3. L'ORF de NS2 est localisé à l'intérieur de l'ORF de NS1 mais leurs codons d'initiation sont chacun dans un cadre de lecture différent et à peu de distance l'un de l'autre (ATG de NS1 à la position 1355, ATG de NS2 à la position 1362). La traduction de NS2 résulte donc vraisemblablement d'un mécanisme «leaky scanning». Le signal de polyadénylation (AATAAAA) localisé en position 3002, soit 17 nucléotides en amont du codon stop de la NS1 est vraisemblablement fonctionnel.

L'ORF4 du PiDENV est le plus grand (2418 nucléotides) et a une capacité de codage de 805 aa. Comme pour les membres du sous-groupe A du genre *Densovirus*, la transcription de cet ORF génère vraisemblablement un seul messager à partir duquel les quatre protéines de capsid (VP1-4) sont synthétisées par le mécanisme de «leaky scanning». Le fait que l'ORF4 commence à seulement 27 nucléotides en aval de la «TATA box» indique que les éléments essentiels du promoteur se trouvent dans les ITRs. Deux domaines conservés sont présents dans cet ORF. Le premier est celui de la protéine VP4 qui joue un rôle essentiel dans l'architecture de la capsid des densovirus et le deuxième est la région N-terminale de la protéine VP1 dans laquelle on retrouve chez quasiment tous les parvovirus deux motifs importants qui sont le motif PLA₂ (phospholipase de type A₂) (DxxAxxHDxxY) et Ca²⁺-binding loop (GPGN). Le motif de PLA₂ est très conservé et joue un rôle essentiel dans la libération du virus de l'endosome tardif. Le Ca²⁺-binding loop est en position 186 de la protéine VP1 et 17 nucléotides en amont du site catalytique HD (nucléotide 207). En outre, la séquence PGYKYL en amont du motif GPGN est conservée dans PiDENV, GmDENV, MIDENV, MVM, PPV tandis que dans B19, seuls P, G, et Y sont conservés (PGxxYx).

La structure 3D de la sous-unité de la capsid du PiDENV (protéine VP4) a été prédite à l'aide du modelage d'homologie Swiss-Model sur la base des fortes similarités (81%) existant entre les séquences des protéines VP4 du GmDENV et du PiDENV. La fiabilité de la structure a été estimée par «QMEAN4 global scores». Le modèle a un score brut de QMEAN4 très bas (0.308) et un score de QMEAN Z très négative (-7.76), c'est-à-dire une fiabilité très basse de la structure prédite. À l'aide du programme UCSF Chimera (50), un modèle identique à celui produit par Swiss-Model a été obtenu. Quand ce modèle a été utilisé pour l'assemblage d'une capsid complète (soixante sous-unités de la symétrie T=1) du virus, il est apparu que les boucles extérieures peuvent se relier et s'entrelacer dans les canyons de la capsid pour former une structure cohérente.

En résumé, le génome du virus PiDENV a été entièrement séquencé et a révélé une séquence de 5990 nucléotides. Le génome possède à chaque extrémité une longue répétition inversée (ITR) de 540 nts. Les 120 premiers nucléotides de chaque ITR peuvent se replier pour former une structure en épingle à cheveux suggérant un mécanisme de réplication de type de «rolling circle». Ce mécanisme, commun aux différents parvovirus, est à l'origine des deux orientations alternatives «flip» ou «flop» dans l'épingle. Le génome du PiDENV est ambisense c'est-à-dire, que les ORFs des protéines structurales (VP) et non-structurales (NS) sont chacun

localisé dans la moitié 5' des brins opposés. Sur un brin, 3 ORFs codent potentiellement les protéines NS1, NS2 et NS3. Par convention, les gènes NS sont représentés à gauche du génome viral. Sur le brin complémentaire, un ORF de grande taille occupe toute la moitié 5' et code les protéines structurales VP1, VP2, VP3 et VP4. La séquence nucléotide et l'organisation du génome de PiDENV justifient la classification de ce virus au sein du sous-groupe A du genre *Densovirus* de la sous-famille *Densovirinae* de la famille *Parvoviridae*.



Étudiante



Date



Directeur de recherche



Date

INTRODUCTION

The family *Parvoviridae* includes small (around 25 nm in diameter), non-enveloped icosahedral viruses whose genome is a single-stranded, linear DNA and about 4 to 6 kb in length. These viruses are classified into two subfamilies: *Parvovirinae* infecting vertebrates and *Densovirinae* infecting invertebrates. Densoviruses (DNVs) comprise viruses isolated from arthropods, mainly insects and cause high mortality in infected hosts. In addition to DNVs and all vertebrate parvoviruses with monosense genomic organization (structural and non-structural proteins encoding sequences present on the same strand), there are ambisense DNVs of which two representatives are *GmDENV* and *JcDENV*. All parvoviruses possess palindromic sequences at the extremities of the viral genome. Although these terminal sequences are dissimilar in length and in structure between different parvoviruses, they proved to play an important role in viral DNA replication. To date, the near-atomic structures of the three densoviruses (*BmDENV*, *PstDENV* and *GmDENV*) have been solved displaying a capsid composed of 60 subunits of identical proteins some of which include a domain with phospholipase A₂ (PLA₂) activity.

Chao *et al.* (1985) (13) isolated a densovirus, PiDENV, from larvae of soybean loopers (*Pseudoplusia includens*). Up until now, only the physicochemical and serological characteristics of this virus were resolved (13). Because PiDENV can be used as a biological control against soybean loopers, an important agriculture pest, we would like to determine further molecular characteristics of this virus. The four objectives of the work were the following:

1. Cloning and sequencing the genome of PiDENV. In this objective, we would like to have a clone with the whole viral genome, including the complete inverted terminal repeats (ITRs) at two extremities of the genome in a plasmid vector. Then the complete nucleotide sequence of PiDENV genome will be determined by sequencing.
2. Determining the organization of the genome of PiDENV from which expression strategy of the virus can be predicted.
3. Comparing the genomic organization of PiDENV with other densoviruses.
4. Prediction of the 3D structure of PiDENV capsid subunit and capsid.

REVIEW OF LITERATURE

1. The family *Parvoviridae*

1.1. Classification

From the first parvovirus isolation (Kilham rat virus) (39) in 1959, the family *Parvoviridae* has been rapidly expanding to include all viruses that are small, isometric, non-enveloped and have linear single-stranded DNA. This family is divided into two subfamilies based on host range: *Parvovirinae* which infect vertebrates and *Densovirinae* which infect invertebrates. Phylogenetic analysis and molecular characterization allows further classification of members within a subfamily into specific genera and species. Up until now, there are five genera in the subfamily *Parvovirinae* and four genera within the subfamily *Densovirinae*. These genera involve 37 accepted species and many unassigned or tentatively assigned to a genus or species (65).

Although also isolated from Crustacea, most of densoviruses (DNVs) are from insects including Lepidoptera, Diptera, Orthoptera, Dictyoptera and Odonata (66). According to the phylogeny of the most conserved non-structural protein sequences (Figure 1), the individual species within the subfamily differ by at least 5 percent. In addition, natural infection of each species seems usually to be confined to single host species. Only half of around 30 reported DNV isolates are presently classified due to the lack of sufficient genomic characterization. In general, DNVs were classified to four genera: *Densovirus*, *Iteravirus*, *Brevidensovirus* and *Pefudensovirus*. Unlike iteraviruses and brevidensoviruses with monosense genome organization in the *Densovirinae* subfamily, genomes of members in the *Densovirus* and *Pefudensovirus* genera are ambisense, meaning that their non-structural (NS) proteins are encoded on the left half of the genomic DNA strand and structural (VP) proteins on the right half of the complementary strand of the single-stranded DNA genome. Upon infection a double-stranded genome is created by self-priming from its terminal hairpin so that both gene cassettes are present on the same molecule. The ambisense densoviruses are further subdivided into 3 subgroups: subgroup A, B and C. Subgroup A of the genus *Densovirus* is composed of viruses infecting Lepidoptera (butterflies and moths), represented by *Junonia coenia* DNV (*JcDNV*) (24) and *Galleria mellonella* DNV (*GmDNV*) while two other groups encompass viruses isolated from different orders of insects (68). In spite the fact that all viruses in the genus have ambisense organization, viruses in subgroup A display high sequence identity (80-90 percent)

while those in subgroup B show very low sequence identities (10-30 percent) maybe because they adapted to hosts of different insect orders (66).

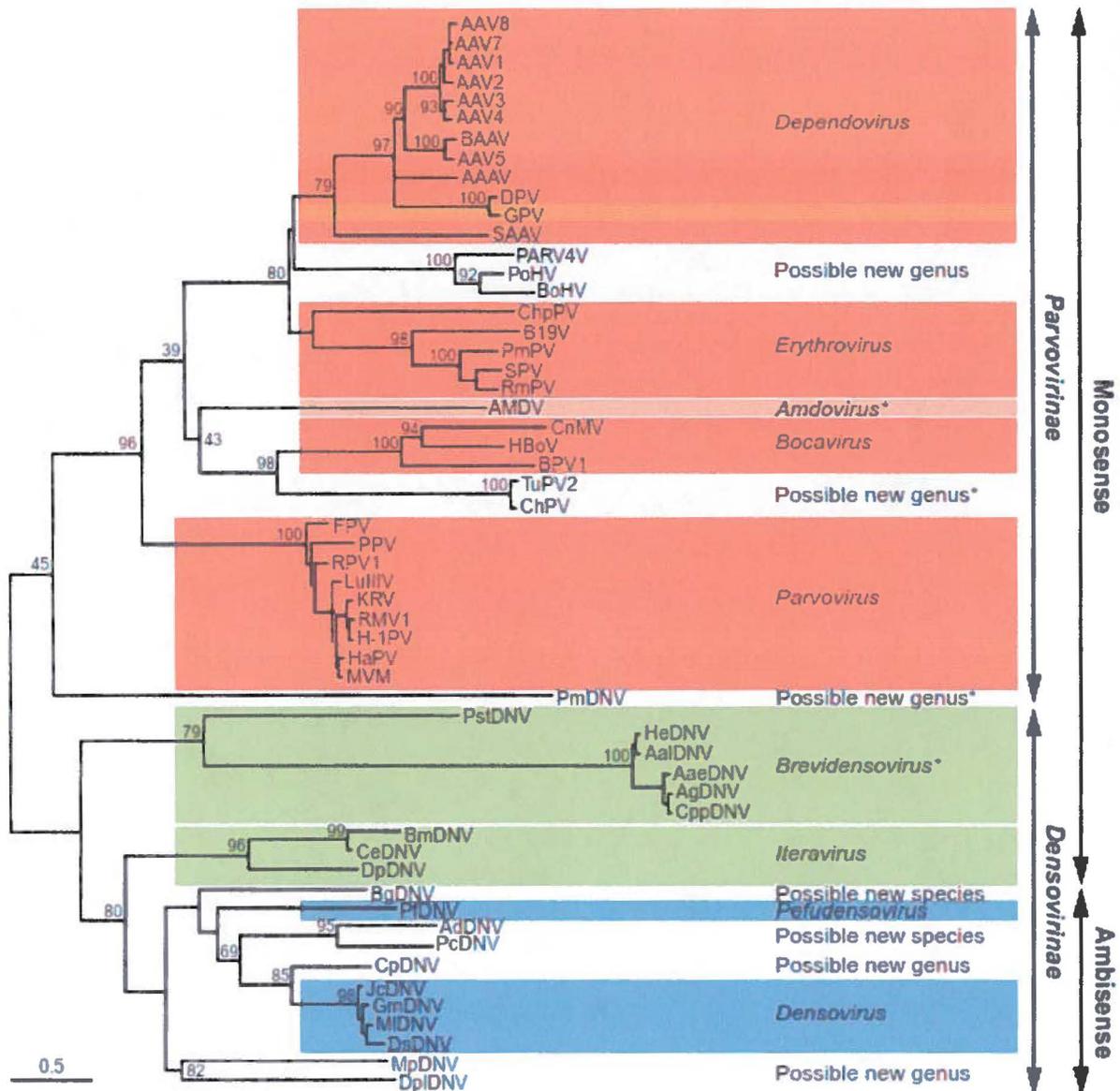


Figure 1. Phylogenetic relationship among the pleiotropic NS1 proteins of members of the family Parvoviridae. The tree was constructed with the programs included in the Phylip package at the <http://mobyle.pasteur.fr/cgi-bin/portal.py> website (ClustalW-multialign, Phylip distance matrix, PROTDIST, Neighbor 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. The scale bar represents the rate of amino acid substitutions. This phylogenetic analysis distinguishes the two subfamilies, as well as the monosense and ambisense densovirus, and recognizes the different genera as separate clades. Several other clades that are possible new genera are also recognized. The asterisk indicates clades of viruses that do not contain the phospholipase A₂ motif in their capsid proteins. Reprinted from reference (68).

1.2. Virus particle

1.2.1. Viral capsid

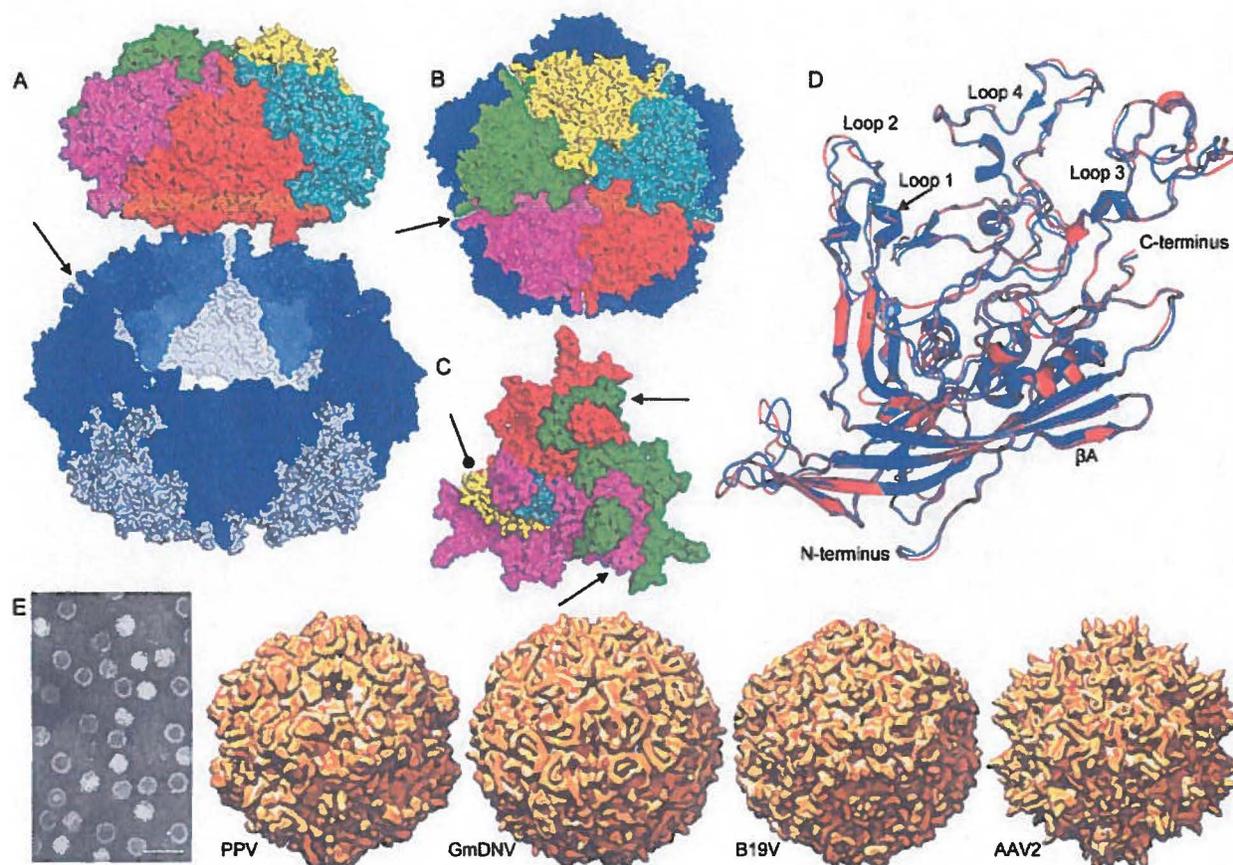


Figure 2. Morphology of parvoviruses. (A) Side-view, at a resolution of 3.5 Å, of a tilted model of a porcine parvovirus (PPV) capsid with the top five trimers translated 120 Å along its 5-fold axis from the middle body of 10 trimers (five shown in dark blue and five in light blue); the bottom five trimers are not shown. The 5-fold axes are located at the intersection of five trimers (e.g. at arrow: green, magenta, two dark blue and one light blue colored trimer). (B) Top view of model shown in (A), without light-blue trimers, along the 5-fold axis. The channels at the 5-fold axes are clearly visible (arrow indicates same 5-fold axis as shown by arrow in (A)). (C) Structure of a PPV trimer. The arrows indicate the intertwining of the GH-loop (between β strands G and H) in the counter-clockwise located proteins from near the 3-fold axis towards the 2-fold axis (arrows). The GH-loop actually consists of two loops of which one (loop 3, yellow), running to the 2-fold axis, is partly covered by loop 4 (cyan) near the 3-fold axis (oval arrow). (D) Parvoviruses may have remarkably similar structures despite low sequence identities. This figure shows the alignment of minute virus of mice (MVM, in red; 549 amino acid residues, PDB 1mvm) and PPV (in blue; 542 amino acid residues, PDB 1k3v) structural proteins that have only 52% sequence identity. Nevertheless, 528 CV's (97%) of the residue pairs occupy the same position in the capsid (root-mean-square error of 1.0 Å). (E) Negative contrast electron micrograph of empty and full PPV particles (bar = 50 nm) and space-filling models of the capsid structures of PPV, *Galleria mellonella* densovirus (GmDENV; PDB: 1DNV), human B19 virus (B19V; PDB: 1s58), and adenoassociated virus-2 (AAV-2; PDB: 1lp3) shown at a resolution of 4 Å. In each case, the view is down a 2-fold axis at the centre of the virus, with 3-fold axes left and right of centre, and 5-fold axes above and below. Models (A-D) have been rendered by PyMOL and the space-filling models by CHIMERA (Multiscale Models). Reprinted from (68).

Parvoviral virions are very compact (“parvo” meaning “small”). They are 215-255 Å in diameter, non-enveloped and icosahedral (six 5-folds, ten 3-folds and fifteen 2-folds) with T=1 symmetry (Figure 2). Some parvoviruses (GmDENV) have smooth structures, some (AAV-2, PstDENV) have spiky ones (3- or 5-fold symmetry axes). In general, parvoviral particles are stable and can survive at room temperature for months or years with little loss of infectivity. The rugged virion is made up of structural proteins VP1-4 to protect the vulnerable DNA genome inside. Its molecular mass is about $5.5\text{-}6.2 \times 10^6$, buoyant density in aqueous isopycnic cesium chloride gradient is $1.39\text{-}1.43 \text{ g/cm}^3$, and sedimentation coefficients (S_{20w}) is approximately 110-120 S. While spermidine, spermine and putrescine were found in some insect viruses (GmDENV, JcDENV and BmDENV-1), cellular enzymes or chromatin components were not found in parvoviral particles. All parvoviruses, except brevidensoviruses, amdovirus, chicken and turkey parvoviruses and PmDENV, have a PLA₂ domain in the minor structural protein which can play important role in virus entry and infection (68).

1.2.2. Viral genome

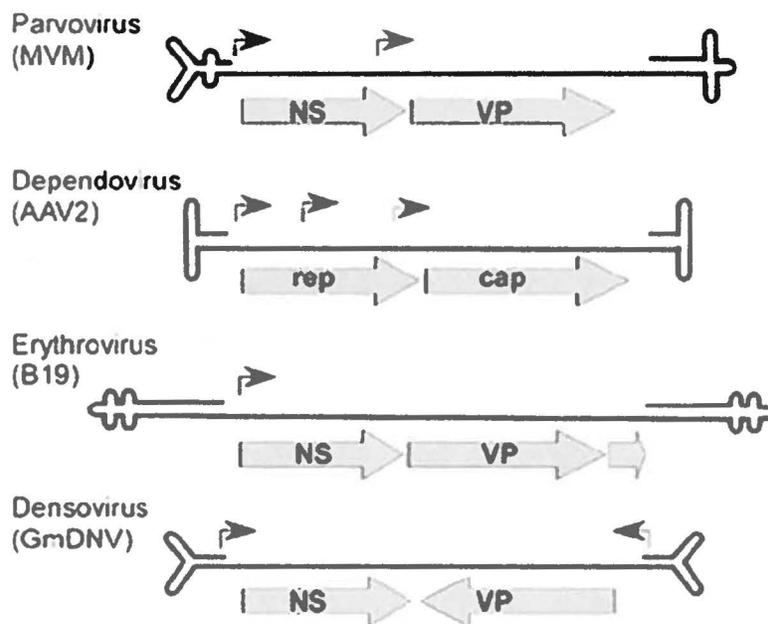


Figure 3. Genome organization of parvoviruses. Genomes of representatives of four out of the nine genera of *Parvoviridae*. Their terminal hairpin structures are magnified around 10-fold of the single-stranded region. The promoters are indicated by arrows, major gene blocks by open arrows with N- to C- terminal direction. Reprinted from (18).

As a result of the tiny size of the viral particle which may confer the virus the capacity to enter the nucleus of host cell, only a limited size of the parvoviral genomes can be tolerated. In fact, they are exclusively 4-6 kb in length and comprise roughly 25 percent ($1.5-2 \times 10^6$) of the particle mass. These genomes are single-stranded linear DNA which terminate in short palindromic sequences at two ends (Figure 3) (18). Within a viral particle, either positive- or negative-sense DNA can be packaged. The ratio of these two kinds of particles depends on the specific virus, either equal polarity or predominantly minus. All parvoviruses possess two gene cassettes which are, by convention, non-structural proteins on the left hand and structural proteins on the right hand. In addition to overlapping reading frames, parvoviruses also have transcriptional promoters and splicing signals integrated in the same primary sequence. Furthermore, many parvoviral genomes are highly repetitive maybe due to the unidirectional, strand-displacement process of their genome replication. These repeats have their own benefits, for example, they can be binding sites for NS1 proteins (5'-ACCA-3' motif) (18).

1.2.2.1. Telomeres

The telomeres are the hallmark of the *Parvoviridae* family. These imperfect palindromes of about 120-550 nucleotides are short imperfect palindromes. These sequences are able to fold back on themselves to create duplex secondary structures that vary in size, sequence and shape between genera or even between species in the genus *Dependovirus*. The terminal hairpins can be different in some viruses, such as left and right hairpins of Minute Virus of Mice (MVM), while in others they are part of inverted repeats at two extremities so that their sequences are closely related or identical like the hairpins of viruses in the genus *Densovirus*. These terminal palindromes can be small internal palindromes forming Y-shape, for example the hairpins of GmDENV, or T-shape like the left-hand end of AAV. On the other hand, these sequences maybe severely asymmetric or contain one or more clusters of mismatched nucleotides near the axis of the palindrome (18).

Despite the variety in size, sequence and shape, all parvoviral telomeres have certain functions in viral life cycle. First of all, these hairpins contain cis-acting information for the viral genome to be replicated and packaged. Parvoviruses have a special mechanism of DNA genomic replication which is called "rolling-hairpin replication" (RHR). In this process, parvoviruses use the terminal hairpins as primers for synthesizing the complementary strand.

After monomer or replicative intermediates were formed, viral replication origin within the hairpin is exposed to allow a protein (NS1 or Rep68/78) to nick the DNA to separate unit-length genomes and also allow the termini to be copied. Resulting from this process, an inversion of the terminus occurred when each round of replication ends. As all parvoviral hairpins are imperfect, the synthesized hairpins exhibit two configurations, which are named “flip” and its reverse-complementary version “flop”. In the genome replication, these palindromes can also act as hinges where the direction of replication is reversed at the end of each unit-length genome. This ability seems to depend on the imperfect or base-paired regions around the axis of each terminus which can provide favourable environment in energy for the hairpins to unfold and refold. Moreover, transcriptional control elements for nearby viral promoters were found in telomere sequences. Besides, viral telomeres contain sequences or structures that promote interactions between viral genome and capsid in viral packaging. Viruses with identical telomeres package their positive- or negative-sense DNA genome with equal ratio whereas those with different telomeres usually package their genomes disproportionately (except LuIII virus) (18).

1.2.2.2. Genome organization of subgroup A ambisense densoviruses

Genomes of viruses in this group are about 6 kb in length including two long inverted terminal repeats (ITRs) which are approximately 550 nucleotides, making up around 18 percent of the genome (Figure 4). About 130 nucleotides at the extremities can fold back into themselves to form hairpins with typical Y-shapes (Figure 5). Additionally, flip/flop configurations are also found at both ends of the genomes (66). The remaining sequence of ITRs contains TATA boxes and upstream promoter elements for NS and VP genes (67).

Unlike vertebrate parvoviruses, ambisense densoviruses have three different NS proteins (NS1-3, with molecular mass 20, 30 and 60 kDa) which are encoded in similar ways among viruses within the group. In fact, the 5'-half of open reading frame (ORF) of NS1 overlaps the ORF of NS2 although the proteins are encoded in different reading frames. Just upstream of the NS1/2 cassette is NS3 ORF (66). Transcript mapping of GmDENV (67) and MIDNV (25) revealed the expression strategy employed by these ambisense densoviruses. The 20-kDa NS3 protein is translated from a 2.5 kb transcript initiated at the upstream promoter within the 5'ITR while NS1 and NS2 proteins are translated from a 1.8 kb transcript after splicing out the NS3 coding sequence from the initial 2.5 kb transcript. The translation of NS2

proteins appears to be done via leaky scanning mechanism because the upstream initiation codon of NS1 is in a poor context (YnnAUG). In addition, the proximity of the initiation codon of NS1 and that of NS2 possibly promotes alternative initiation. Ding *et al.* (22) showed that NS1 of JcDENV recognizes and binds the (GAC)₄ sequence in the stem of the hairpin. This enzyme subsequently nicks the single-stranded forms of the hairpin preferentially at two sites (G*TAT*TG). It was demonstrated that this protein possesses helicase activity with less conserved Walker A-motif (GxxxxGK [T/S]). Another non-structural protein, the NS3 of JcDENV contains two zinger-finger motifs, and was shown to be essential for viral replication (1). NS1 is the most conserved protein among ambisense densoviruses whereas no highly conserved domains were found in NS2 or NS3.

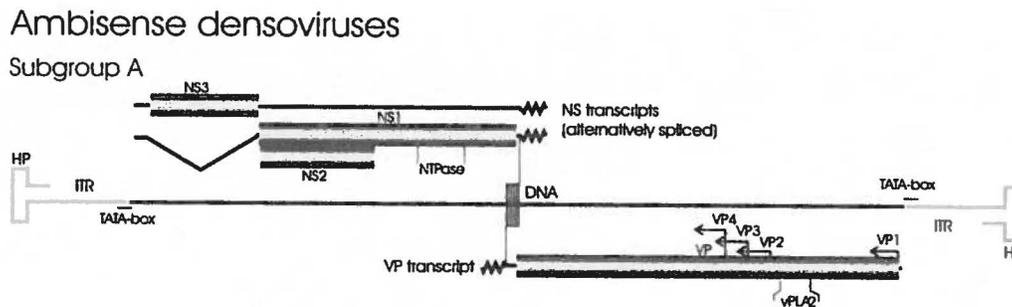


Figure 4. Genome organization of ambisense densoviruses. All viruses have ITRs and their structural proteins have a PLA₂ motif. The large NS proteins contain NTPase (ATPase) and rolling circle replication (RCR) motifs (66).

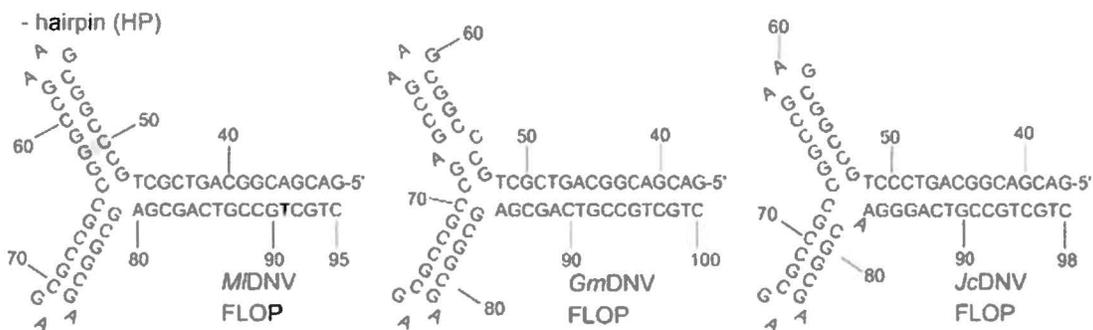


Figure 5. Terminal hairpin of subgroup A densoviruses. All of them have Y-shape and each have two configurations: flip and flop (66), e.g. nucleotides 48-79 in MIDNV have the reverse-complement sequence in the flip configuration.

In contrast to non-structural proteins, structural proteins of subgroup A densovirus are encoded by only one ORF. After the VP mRNA is transcribed, the four VP proteins (45, 53, 58 and 89 kDa) are generated by a leaky scanning mechanism (25, 67). The two most important

conserved motifs were detected in all ambisense densoviruses as well as in iteraviruses and vertebrate parvoviruses (81). These are the Ca^{2+} -binding loop (GPGN) and PLA_2 (DxxAxxHDxxY) motifs, even though the specific activity of densovirus PLA_2 is significantly lower than that of iteravirus or vertebrate parvoviruses (25, 26, 42, 67, 81). Similar to other PLA_2 s, the distance between the Ca^{2+} -binding loop and the active site (HD) varies slightly in the densoviruses. Additionally, all the PLA_2 identified up until now have been on the least abundant capsid protein VP1 (66).

1.3. Viral life cycle

As tiny particles and thus with limited genetic capacity, parvoviruses must equip themselves with numerous mechanisms to take advantage of the host cell throughout the whole virus life cycle from entry to egress (77). The earliest step in virus infection is attachment to the host cell, which is mediated by interactions between viral capsid molecules and receptors and/or coreceptors on host cell surface. The parvovirus capsid proteins involve a highly conserved β -barrel motif and hypervariable regions connecting the eight strands of the β -barrel. These variable regions may be responsible for usage of cellular receptors which may be heparan sulfate proteoglycan (AAV2) (63), sialic acid (AAV4, AAV5, AMDV, BPV, MVM) (17, 75), erythrocyte P antigen globoside (B19) (8) or transferrin (CPV, FPV) (49). Following attachment, entry of parvoviruses to host cell is thought to be accomplished via receptor-mediated endocytosis. However, how these viruses escape endosomes remain unclear (77). Moreover, there are multiple pathways of intercellular trafficking in different cell types and hence the specific way employed depends on the distinct virus and its host cell. Even though parvoviruses have to bring their genomes to the nucleus for replication, the nuclear transport mechanism is still not known. Nevertheless, viral capsid appears to play important role in translocation of viral particle to the nucleus as well as in uncoating of its genome. Specifically, the VP1 protein with PLA_2 activity on its unique N-terminus is required for these events (30, 81).

Once released into the nucleus, DNA of autonomous parvoviruses wait for the cell to enter S-phase to promote the conversion of single-stranded DNA molecule into double-stranded DNA. This DNA can then be a template for transcription and subsequent replication. The first genes to be transcribed are non-structural genes (NS1, NS2 and NS3 for densoviruses, NS1 and NS2 for autonomous parvoviruses, and Rep52, Rep68 and Rep78 for

dependoparvoviruses (AAV)) which are required for viral DNA replication. After these gene products accumulate, viral DNA amplification begins (77). Parvoviruses exert a novel mechanism of replication of their genomes, which is called “Rolling-hairpin replication” (RHR) as depicted in Figure 6 (19). This unidirectional, strand-displacement process is unique in using the short imperfect palindromic sequences at the ends of the linear single-stranded DNA as replication origin, as well as self-primers, and also in the creation of series of duplex intermediate replicative forms. As presented in step (i) of Figure 6, the left hairpin of the negative-strand DNA primes the synthesis of the complementary strand to yield the first duplex intermediate (ii). NS1 protein then comes to nick the covalently-closed right hairpin (iii) and attaches covalently to the 5' end at the nick. Next, the replication fork unfolds and copies the right hairpin so that the original sequence (R) is replaced by its inverted complement (r), as shown in step (iv) of Figure 6. Because terminal hairpins are imperfect and this inversion occurs at every replication round, progeny genomes have two configurations, designated “flip” and its reverse-complement “flop”. This process is called “terminal resolution” which is used generally by parvoviruses with identical telomeres including group A ambisense densovirus, AAVs and B19. In contrast, viruses with unlike termini, e.g. MVM, replicate their left telomere by terminal resolution and their right telomere by using “junction resolution” which conserves a single sequence orientation and activates NS1 in different manner. At the end of step (iv), extended form duplex termini are created which are then melted out causing single strand to fold back on themselves to create “rabbit ear” hairpins (v). This formation allows strand-switching and self-priming to synthesize additional sequences (vi), giving rise to duplex dimer (vii). Step (viii) is repeat of step (v) and after another extension, concatemers of viral genome are obtained. NS1 can finally excise the multimer to produce monomers of viral genome, some of which will have flip, some will have flop orientation. These genomes are ready to be packaged to generate new progeny virions. This step in viral life cycle happens in the nucleus and viral ITRs appears to play an important role in encapsidation (76, 79).

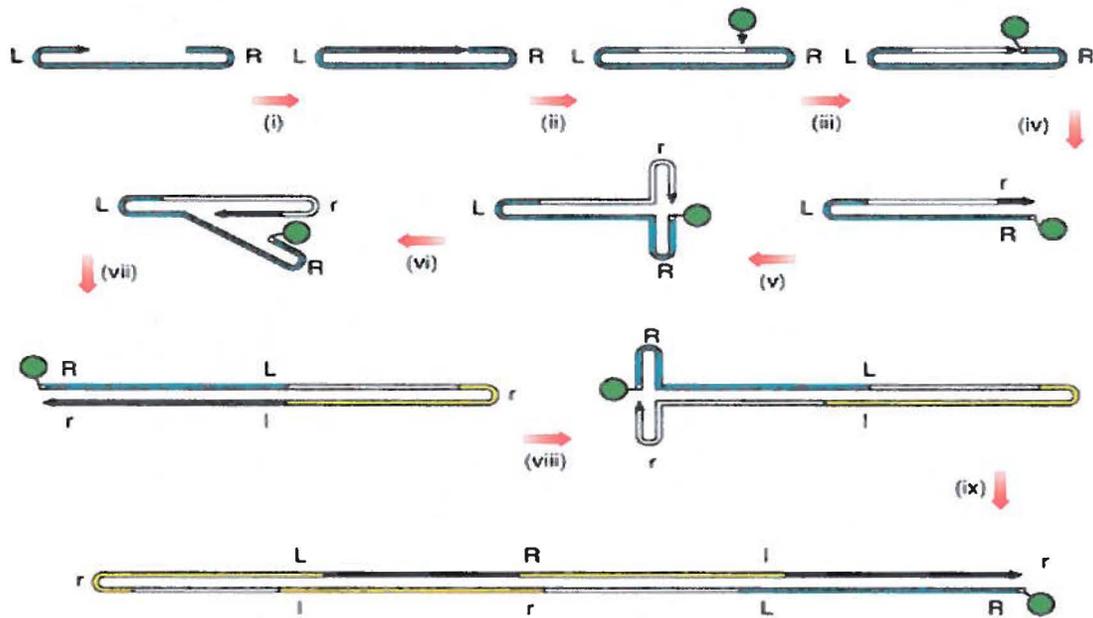


Figure 6. Parvoviral "rolling-hairpin" DNA replication. The parvoviral genome is represented by continuous line, blue for original genome, yellow for progeny genome. Black bar with an arrow at its 3' end is newly synthesized DNA. NS1 is represented by green circle. L and R are palindromic sequences at each terminus and its reverse complement are depicted by l and r, respectively (19).

2. *Pseudoplusia includens* densovirus (PiDNV)

Chao *et al.* (13) isolated two full virus-like particles from field-collected *Pseudoplusia includens* larvae (also called: soybean looper or *Chrysodeixis includens* (Hübner) (Noctuidae, Plusiinae, Lepidoptera)). Under the electron microscope, their sizes were determined, one is 20 nm and the other is 25 nm. When injected separately to *Pseudoplusia includens* larvae, each caused high mortality for its hosts. The 25 nm particles were later identified as an insect picornavirus while the 20 nm particles were believed to be densovirus. The latter particles were thus named *Pseudoplusia includens* densovirus (PiDNV) and their physicochemical and serological characteristics were determined.

PiDNV particles, as shown by electron microscopy, are icosahedral (13). They have a sedimentation coefficient of about 120 S and a buoyant density of 1.40 g/cm³. Absorption spectrum indicates that the viral particle is composed of nucleoprotein. Furthermore, positive diphenylamine reaction and negative orcinol reaction demonstrate that PiDNV includes only DNA, making up 37.8% of the total particle. Additionally, acridine orange staining as well as reaction to formaldehyde indicate that viral DNA is single-stranded. Agarose gel electrophoresis gave one sharp band on the gel which is about 6 kb in length (Figure 7C). Interestingly, inverted

terminal repeats (ITRs) were observed as panhandle-like structure (Figure 7A and 7B) by electron microscope and consisted of 6-7 % of the genome. SDS-polyacrylamide gel electrophoresis showed four viral proteins (VP1-4) with molecular weights of 46.5, 54, 64 and 87×10^3 . Finally, immunodiffusion indicated that antiserum of GmDENV reacted with both GmDENV antigen and that of PiDENV but they are not identical while antiserum of PiDENV reacted only with PiDENV antigen (13).

Soybean looper, the host of PiDENV, can be found in Canada, Mexico and the United States and can feed not only on soybean but also on a wide range of important cultivated plants such as sweet potato, peanut, cotton, tomato, crucifers, etc. (12). PiDENV can thus be used as a biological control against these harmful pests. However, since the first characterization of this virus by Chao *et al.* (13), molecular characteristics of PiDENV remained to be unsolved. Here we reported the characterization of PiDENV genome with an analysis on its organization and expression strategy.

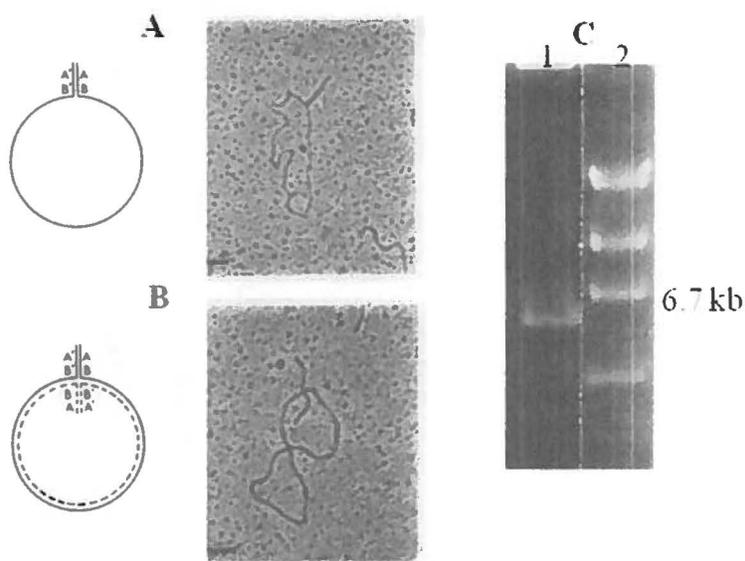


Figure 7. DNA of PiDENV. A. Single-stranded DNA. B. Double-stranded DNA. Both DNA in A and B have panhandle-like structure at two extremities, an evidence of ITRs. C. Agarose gel electrophoresis of PiDENV DNA with a band of around 6kb on lane 1 and lane 2 is a DNA marker (13).

METHODOLOGY

1. Cloning and sequencing the genome of PiDNV

1.1. Isolation and purification of PiDNV DNA from infected larvae

The infected soybean loopers larvae were provided by Dr. Yu-Chan Chao, University of Arkansas, in 2000. They have been stored at -20°C since then. To isolate viral DNA, one infected larva was ground with a pestle in a mortar. 500 μl of sterile H_2O was then added and the grinding was continued until a homogeneous suspension was achieved. Next, the suspension was transferred to an Eppendorf tube, 500 μl of sterile H_2O was used to wash the mortar and pestle and all the mixture was transferred to the same tube prior to be centrifuged at 3000 rpm for 10 minutes. The supernatant, which contains the virus, was transferred to a new tube. 200 μl of this suspension was used to extract viral DNA by mixing with 300 μl of lysis buffer (6 M guanidine-HCl, 10 mM Tris-HCl, 20% Triton X-100, pH 4.4, 80 $\mu\text{g}/\text{mL}$ poly(A) carrier RNA and 0.2 mg/mL proteinase K). After 10 minutes of incubation at 70°C , 125 μl of isopropanol was added and vortexed. The solution was then transferred to a spin column (Qiagen) and centrifuged at 6000 rpm for 2 minutes. The flow-through was discarded and the column was washed twice with 70% ethanol. The column was then centrifuged at 9000 rpm for 1 minute to remove all traces of ethanol before 50 μl of H_2O was added to the center of the membrane in the column. 40 μl of viral DNA-containing solution was collected after centrifugation at 9000 rpm for 2 minutes. This was finally stored at -20°C for downstream applications.

1.2. Digestion of PiDNV DNA with the ClaI enzyme

500 ng of the genomic DNA of PiDNV was digested with 1 u of ClaI (Invitrogen) in the presence of 1X buffer M (10 mM Tris-HCl, pH 7.5; 10 mM MgCl_2 ; 1 mM dithiothreitol; 50 mM NaCl) at 37°C for 2h. The digest was then run on agarose gel electrophoresis to see if the enzyme can cut the DNA.

1.3. Agarose gel electrophoresis

In general, 0.8% agarose gel was prepared in buffer TAE (40 mM Tris base; 1 mM ethylenediaminetetraacetic acid (EDTA), disodium salt, pH 8.0 and 0.114 % glacial acetic

acid). The DNA was loaded into a well on the gel by using a loading buffer (0.25% of bromophenol blue, 0.25% of xylene cyanol FF, 30% of glycerol). To estimate the size of the DNA of interest, a DNA ladder (1kb or 100bp, Gibco®BRL) was also used. 100 V was then applied for 45 minutes using the Biorad agarose electrophoresis apparatus. After electrophoresis, the gel was stained with ethidium bromide and the DNA bands were visualized under an UV transilluminator.

1.4. Preparation of insert

The viral DNA (2 µg) was first blunt-ended with 2 u of Klenow (the large fragment of DNA polymerase I) (New England Biolabs, NEB) in the presence of 100 µM dNTPs (Invitrogen) and 1X buffer 2 (NEB; 50 mM NaCl, 10 mM Tris-HCl, 10 mM MgCl₂, 1 mM dithiothreitol, pH 7.9 at 25°C) at 25°C for 15 minutes.

After that, ethanol precipitation was carried out. 2-2.5 volume of ice-cold ethanol 100% and 1/10 volume of 3M sodium acetate was added to the Klenow reaction and the mixture was let stand for 20 minutes at -20°C before being centrifuged at 15 000 rpm for 10 minutes. The supernatant was then decanted. After being washed with ethanol 70%, the pellet was air-dried for around 1 hour and then DNA was suspended in 20 µl of sterile H₂O.

For cloning of two halves of PiDENV genome, the blunt-ended genome was digested with 5 units of ClaI (Invitrogen) at 37°C for 5 hours. Next, the digest was run on agarose gel electrophoresis to separate digested fragments following the same protocol as described above except that only 75 V and 90 minutes were used. The distinct bands, around 3.2 kb and 2.8 kb in case of ClaI-digested PiDENV DNA, were excised from the gel and placed in separate tubes.

To retrieve the DNA from gel, 300 µl of the solution of Minipreps Express Matrix™ (Qbiogene) was added to one volume of the gel (100 mg) and incubated at 65°C for 5 minutes. The DNA was then precipitated using ethanol as described above.

1.5. Preparation of vector

The circular vector pBluescript KS (+) (Figure 8) was used to clone separately two halves of PiDENV genome. This 2.9 kb vector is a high-copy plasmid and has the ampicillin

resistance gene, lacZ (blue/white screening) and a multiple cloning site including ClaI (5'-overhang) and SmaI (blunt enzyme, 5'-CCC^GGG-3'). To prepare for cloning, 2 µg of the plasmid was digested simultaneously with the two enzymes (Invitrogen), ClaI (5 units) and SmaI (5 units) in 1 X buffer React 4 (20 mM Tris-HCl, pH 7.4; 5 mM MgCl₂; 50 mM KCl) at 37°C for 8 hours so that the vector had a SmaI-blunt end and a ClaI-sticky end. The digest was then incubated at 65°C for 20 minutes to inactivate the enzymes.

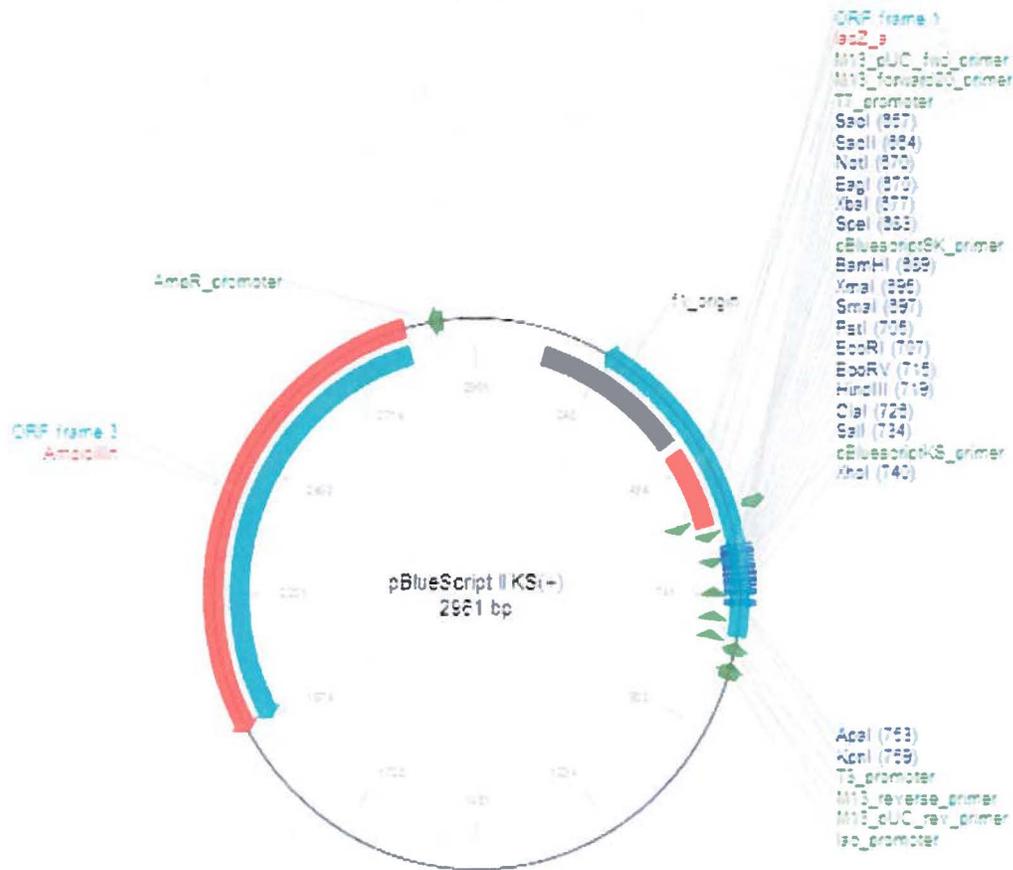


Figure 8. Vector pBluescript KS (+) (Stratagene).

1.6. Ligation

To ligate the ClaI-SmaI digested vector with each half of PiDNV DNA (Klenow and ClaI), a 10 µl reaction was prepared which contains 50 ng of vector, 250 ng of insert, 1 µl of T4 DNA ligase (NEB), 1 X buffer for T4 DNA ligase and sterile H₂O. It is then incubated at 4°C overnight.

1.7. Transformation

1.7.1. Prepare electrocompetent cells

The bacterial cells used are SURE 2 strain (Stratagene). As indicated by its name, SURE (Stop Unwanted Rearrangement Events), the strain lacks the components of pathways that catalyse the rearrangement and deletion of nonstandard secondary and tertiary structures in conventional *E.coli* strains. In fact, they are restriction minus (McrA-, McrCB-, McrF-, Mrr-, HsdR-), endonuclease (*endA*) deficient, and recombination (*recB recJ*) deficient. Because PiDNV genome is expected to contain many secondary structures, especially in the ITRs at the two extremities, we used this strain to ensure that the whole sequence of PiDNV is amplified completely within the bacterial cells. For antibiotic resistance, SURE 2 cells are resistant to kanamycin and tetracycline. Furthermore, they have the *lacIqZAM15* gene, on the F' episome, which allows blue-white screening and thus facilitates the selection of positive clones.

To prepare electrocompetent SURE 2 cells, 2 ml of LB was inoculated with a single colony of SURE 2 in the presence of 10 µg/ml tetracycline and 50 µl/ml kanamycin and cultured overnight at 30°C, with shaking at 200 rpm. After that, the 2 ml overnight culture was used to inoculate 250 ml LB with 10 µg/ml tetracycline and 50 µl/ml kanamycin in an Erlenmeyer flask. This culture was incubated at 30°C with shaking at 300 rpm up until its OD₆₀₀ was around 0.6. It was then incubated on ice for 30 minutes before being centrifuged at 2500 rpm for 20 minutes at 4°C. The pellets were subsequently washed twice with 250 ml of ice-cold sterile H₂O. Next, 10 ml of ice-cold 10% sterile glycerol was added, the mixture was mixed gently and centrifuged at 2500 rpm at 4°C for 20 minutes. 1 ml of sterile GYT (10% glycerol; 0.125% yeast extract; 0.25% tryptone) was added to resuspend gently the pellet, and the competent bacteria were aliquoted to Eppendorf tubes and stored at -80°C. The efficiency as tested with pBluescript KS (+) is 7×10^7 cfu/µg DNA.

1.7.2. Electroporation

The recombinant plasmid was introduced into SURE 2 cells by electroporation. In addition, controls, which comprise (i) only bacteria and (ii) only vector (0.1 ng), were also included in the experiment. Firstly 50 µl of electrocompetent cells were thawed on ice and 1 µl of ligation mixture was added. This mix was transferred to a cold cuvette (BTX disposable

cuvette, 1 mm) and the cuvette was then placed in an electroporator (BXT ECM 395 Electroporation System). 1.8 kVolts was applied for 5 milliseconds. The mixture was then transferred to a tube containing 950 μ l of SOC (20 mM glucose; 20% tryptone; 2% tryptone; 0.5% yeast extract; 0.05% NaCl; 10 mM MgCl₂ and 10 mM MgSO₄), and was cultured at 30°C (200 rpm) for an hour to recover the ability of the recombinant plasmid to resist antibiotics. Subsequently, 100 μ l or 200 μ l of the SOC culture was spread on a LB agar plate containing 50 μ g/ml ampicillin (recombinant pBluescript plasmids are resistant to ampicillin), 80 μ g/ml of X-gal (5-bromo-4-chloro-3-indolyl- β -D-galactopyranoside) and 20 mM of IPTG (isopropyl-1-thio- β -D-galactopyranoside) (for blue/white screening). The plate was incubated overnight at 30°C.

1.8. Plasmid DNA extraction

Single white colonies were cultured overnight (30°C, 200 rpm) in LB-Ampicillin (50 μ g/ml). Plasmid DNA was then extracted from the culture by minipreps. In order to do so, 1 ml of overnight culture was transferred to an Eppendorf tube which was centrifuged at 15 000 rpm for 1 minute. The pellet was mixed vigorously with 100 μ l of ice-cold solution I (50 mM glucose; 25 mM Tris-HCl; 10 mM EDTA; 10 μ g/ml RNase A) before 200 μ l of solution II (0.2 N NaOH; 1% SDS) was added and the suspension was mixed thoroughly. Next, 150 μ l of ice-cold solution III (5 M potassium acetate; 11.5% glacial acetic acid) was added, the suspension was mixed and incubated on ice for 10 minutes before being centrifuged at 15 000 x g for 20 minutes. The supernatant containing the DNA was transferred to another tube to which 0.8 volume of isopropanol was added. After mixing, the tube was centrifuged at 13 000 rpm for 15 minutes, the supernatant was decanted and the pellet was washed with 70% ethanol. Then the pellet was air-dried for about an hour before the DNA was suspended in 30 μ l of sterile H₂O. Five μ l of the extracted DNA were run on agarose gel electrophoresis, as mentioned above, so that its size can be compared with that of a vector and hence the plasmid with an appropriate insert and can be differentiated from an empty plasmid.

1.9. Analysis of clones by restriction digestion

To confirm that the recombinant plasmid contains the proper insert, a restriction digestion was performed. In the case of the clones of two halves of PiDNV genome, Xba I (5'-T[^]CTAGA-3') was used. The restriction site of this enzyme is present on both the vector and the

insert. After cutting with this enzyme, the clone with one half of the PiDENV genome would produce 2 bands of 1.5 kb and 4.6 kb, and the clone with the other half would generate 1.1 kb and 4.6 kb bands on an agarose gel. The restriction digestion was composed of 7 μ l of minipreps DNA; 5 units of XbaI (Invitrogen); 1 X buffer React 2 (50 mM Tris-HCl, pH 8.0; 10 mM MgCl₂; 50 mM NaCl). The reactions were incubated at 37°C for 3h, and the entire digest was loaded on a 0.8% agarose gel and electrophoresis was run as described above.

1.10. Sequencing

Table 1. Sequence of primers for sequencing PiDENV genome.	
5'-3'	3'-5'
GTA AACGACGGCCAG	CATTGGGTATATAAACAACCCAGT
CGGTATGTCCA ACTCAAGAATTGC	ACCACAATAGAAGACTAGCTGG
AACCGCACTGTATTGCTGAAG	TGCTATACCTCGACCTTTCTC
GCGCATATATCAGCGATGTC	TTGGCCTTGTTAGCTGAAC
AAGAAAGCCTATTCGGCAGG	AGAAGCTACCACTAATCCTGG
TGTCTCCTCCAAGTGCTGG	CAATTTTACTGATGAAAAGGAAC
TACCAGGATTAGTGGTAGCTTCTG	TGCCAGTAACGTATCCAAAG
TGTTCACTAAACAAGGCCAAC	TAGATGTTTACTCGACGACTGC
GCCATGGTACTTCAGCTATATCAG	AGCGGATAACAATTTACACAGGAA
TCAGACCAGCTCTGAAAGAG	
TCCTCAGGAAATGCCCTATG	
CGCGCGATGTATTAACCCTC	

Recombinant plasmids containing two halves of PiDENV genome were extracted following the protocol mentioned above except that Minipreps ExpressTM Matrix (Qbiogene) was used to purify DNA instead of isopropanol precipitation. Here 400 μ l of Minipreps Express MatrixTM was added to the supernatant containing DNA (after adding 3 solutions) and the tube was spun for 1 minute (DNA binds instantaneously to the silica in the solution). After the supernatant was decanted, 500 μ l of 80% ethanol was added and mixed. The tube was then spun for 10 minutes, the supernatant was removed and the pellet with DNA was air-dried overnight. 60 μ l of sterile H₂O was used to resuspend the DNA. The tube was spun for 1 minute and the supernatant containing the DNA was transferred to another tube.

In general, 10 μ l of purified DNA (50-100 ng/ μ l) and 10 μ l of primer (5 μ M) (IDT) were sent to MCLAB (USA) for sequencing by automated DNA sequencing (the Sanger's method and primer-walking method (57)). Two clones of two halves of PiDENV genome, which

were designated 5'-half clone and 3'-half clone, were sequenced in both directions with the primers listed in Table 1.

1.11. Cloning of the whole genome of PiDNV

In this cloning, we used the kit Lucigen's DNATerminator® End Repair Kit for blunt-ending the viral genome and subsequently the BigEasy® v2.0 Linear Cloning Kit from Lucigen Corp. for cloning the repaired viral genome to the linear vector pJAZZ-OC (Figure 9) and for maintaining the recombinant plasmids in the TSA™ cells. The procedure of end-repair of viral DNA, ligation and transformation were carried out according to the manufacturer's instructions. The transformed cells were grown on YT agar plates containing 12.5 µg/ml chloramphenicol to allow screening for positive clones.

The first 48 single colonies were used to prepare minipreps and confirmation of positive clone was done by digestion with NotI (NEB).

To have more positive clones, 1800 single colonies were used to do colony PCR with paraffin wax which was demonstrated to increase specificity and sensitivity of PCR (32). In so doing, 5 µl of colony suspension (containing 10 single colonies) was transferred to a PCR tube and then melted paraffin wax was added to the tube. This tube was heated in a thermal cycler at 95°C for 2 minutes and subsequently cooled down to room temperature. At the end of this step, DNA is liberated, protease/nucleases were inactivated and the sample is under the solid paraffin layer. Then 20 µl of PCR master mix was added to the tube (1 X Taq buffer, 0.2 mM dNTP, 0.2 µM forward primer, 0.2 µM reverse primer and 1 u Taq DNA polymerase). The primers are 5'-ATGCACCAGTAGTTAGAGCAG-3' (forward) and 5'-TGCTATACCTCGACCTTTCTC-3' (reverse) which give the amplicon of 1 kb. The cycling condition was 95°C/1 minute (1 cycle); 94°C/40 seconds, 58°C/40 seconds, 72°C/ 40 seconds (35 cycles) and 72°C/5 minutes (1 cycle). 5 µl of PCR was loaded on an agarose gel resolved by electrophoresis for visualizing the amplicons.

Colonies from the positive samples (1 kb band on the gel) were cultured and then non-ionic detergent (NID) miniprep plasmid isolation (41) was performed. In this miniprep, 1.5 ml of bacterial culture was centrifuged at 7000 rpm in 1 minute and the supernatant was discarded. A tip with 150 µl of extraction buffer (5% sucrose; 20–50 mM EDTA; 50 mM Tris

pH 8; 0.75 M NH₄Cl; 0.5% Triton X-100; lysozyme 100 µg/ml, and RNase A 25 µg/ml) was used to loosen the pellet and then the extraction buffer in the tip was released to suspend the bacteria. After that, the tube with suspended bacteria was incubated at 65°C for 5 minutes before being centrifuged at 15 000 rpm for 10 minutes. A sterile toothpick was used to remove the pellet and 120 µl of isopropanol was added to the supernatant and mixed. After the centrifugation of 7000 rpm/10 minutes, the pellet was washed with 70% ethanol and then air-dried. The DNA was then suspended in 30 µl of sterile H₂O of which 10 µl was loaded on 0.8% agarose gel and electrophoresis was carried out. Putative positive clones were digested with NotI to confirm the presence of the appropriate insert.

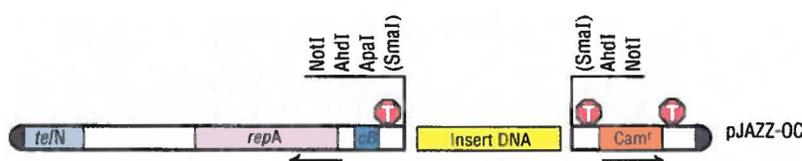


Figure 9. The pJAZZ-OC vector. The left arm is 10 kb and the right arm is 2.2 kb. *Te/N*, protelomerase gene; *repA*, replication factor gene and origin of replication; *Cam*, chloramphenicol resistance gene. Approximate positions of transcriptional terminators are indicated. Reprinted from Lucigen Corp.

1.12. Sequencing of hairpins in linear pJAZZ clones

The 3'-hairpin was sequenced from the linear clones with an internal viral primer (5'-GTCAAGGTCACCGTGACGTCATC-3') and the purification of DNA for sequencing is similar to that described above. However, the 5'-hairpin cannot be sequenced likewise. Instead, two specialised PCRs were used to amplify the hairpin before sequencing. This method was proven to be efficient for amplifying difficult sequences with high GC content (47) thanks to the combination of betaine, DMSO and 7-deaza dGTP. The components of the first PCR were composed of 1 X Taq buffer, 1 unit Taq DNA polymerase, 0.2 mM dNTP (Invitrogen), 0.5 µM forward primer, 0.5 µM reverse primer, 1.3 M betaine, 5% DMSO, 50 µM 7-deaza dGTP (NEB) and around 1 ng of template (pJAZZ-OC clone). The forward primer was vector-specific (5'-ACAACCTTGCTAGGC-3'), the reverse primer was virus-specific (5'-TAGATGTTTACTCGACGACTGC-3') and the product size was around 0.6 kb. The cycling conditions were 94°C/3 minutes (1 cycle); 94°C/20 seconds, 52°C/20 seconds, 72°C/20 seconds (25 cycles) and 72°C/5 minutes (1 cycle). The second PCR was a nested PCR of which the components and cycling conditions were the same as the first PCR except that reverse primer

was an internal viral primer (5'-GTCAAGGTCACCGTGACGTCATC-3') and the product size was about 200 bp. These latter amplicons were then sequenced with the viral internal primer. The preparation of DNA for sequencing was as described above.

2. Bioinformatics

2.1. Primer design

All the primers were designed using the software ApE (<http://biologylabs.utah.edu/jorgensen/wayned/ape/>).

2.2. Assembly of sequence of PiDENV genome

The sequenced contigs were assembled by the CAP3 program (34) (<http://pbil.univ-lyon1.fr/cap3.php/>).

2.3. Structure analysis of viral hairpins

Palindromes in PiDENV sequence were detected using the program einverted (<http://emboss.bioinformatics.nl/cgi-bin/emboss/einverted>). Secondary structures as well as flip/flop configurations of viral hairpins were predicted using the mFOLD program (83) (<http://mfold.rna.albany.edu/?q=mfold/DNA-Folding-Form>).

2.4. Genome sequence analysis

Sequence of PiDENV genome were searched using basic local alignment search tool (BLAST) (82) against the GenBank (nucleotide collection database).

Open reading frames (ORFs) were identified using the ORF Finder of NCBI (<http://www.ncbi.nlm.nih.gov/gorf/gorf.html>) with standard genetic codes. Alignment of amino acid sequences and search for conserved domains were found using Constraint based alignment tool (COBALT) (<http://www.ncbi.nlm.nih.gov/tools/cobalt/cobalt.cgi?CMD=Web>) (48) and Conserved domain database (CDD) (<http://www.ncbi.nlm.nih.gov/cdd/>) (44). Protein motifs were identified by PROSITE (<http://prosite.expasy.org/>) (58) and Motif Scan (http://myhits.isb-sib.ch/cgi-bin/motif_scan) against Prosite, Pfam and HAMAP profiles.

Promoter sequences and transcription factor binding sites were predicted using LBL Promoter program (http://www.fruitfly.org/seq_tools/promoter.html) and TSSG (<http://linux1.softberry.com/berry.phtml?topic=tssg&group=programs&subgroup=promoter>) (55, 60), selected transcription factor binding sites were from tfd.file of Ghosh database (29).

2.5. Prediction of 3D structure of PiDENV capsid subunit and capsid

The 3D structure of GmDENV, a closely related densovirus to PiDENV, had already been obtained by X-ray crystallography. Therefore we used this virus as template. Firstly, the sequence of VP of PiDENV was aligned with the resolved VP sequence (PDB-ID 1DNA, chain A) (59). Then the model was created by homology-modeling using Swiss-Model (automated mode and co-ordinates) (<http://swissmodel.expasy.org/>) (38). After that, Qmean-Z score was calculated to see the reliability of the model and to relate the reliability to segments in the model. A matrix of subunits in the capsid was then created and the compatibility of the model was observed within the matrix. The structures of the subunit and the viral particle were generated by the UCSF Chimera software (50) (<http://www.cgl.ucsf.edu/chimera/>).

RESULTS

1. Cloning of PiDNV genome

1.1. Isolation and purification of PiDNV DNA

About 4 μg of viral DNA was purified from one infected larva. Approximately 200 ng of the DNA was visualized by agarose gel electrophoresis along with the 1kb DNA ladder. The size of DNA of PiDNV genome was around 6 kb as shown in Figure 10.

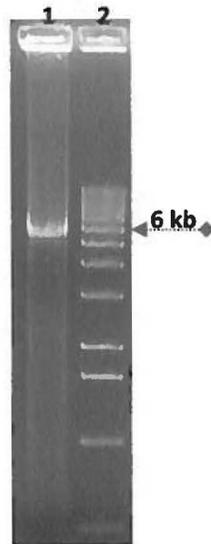


Figure 10. Purified PiDNV DNA. Lane 1: PiDNV. Lane 2: 1kb DNA ladder.

1.2. Cloning of two halves of PiDNV genome

Because it is difficult to clone the whole viral genome, we cut it into two halves and cloned these two halves separately. GmDNV (6039 nucleotides in length), a related densovirus, possesses a *Cla*I site (5'-AT[^]CGAT-3') at nucleotide 4139. Therefore we evaluated whether PiDNV also has this site by digesting its genomic DNA with *Cla*I. Indeed, the enzyme cut the viral DNA into two halves which are about 3.2 kb and 2.8 kb in length. The 3.2 kb-half was designated 5'-half and the 2.8 kb-half was designated 3'-half. These two *Cla*I-digested halves were then cloned separately into the vector pBluescript KS (+) (2.9 kb, *Cla*I-and-*Sma*I-digested). In this cloning, we got 2 clones for 5'-half (6.1 kb) and 2 clones for 3'-half (5.7 kb). The positive clones were digested with *Xba*I to confirm appropriate insertions. As presented in

Figure 11, 3'-half clones (lane 1 and 2) gave, as expected, 2 bands of around 1.1 kb and 4.6 kb while 5'-half clones (lane 3 and 4) gave 2 bands of 1.5 kb and 4.6 kb.

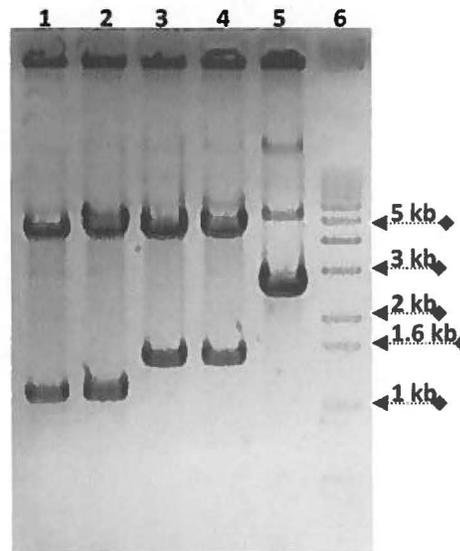


Figure 11. Clones of two halves of PiDnV genome (digested with XbaI). Lane 1 and 2: 3'-half clones; Lane 3 and 4: 5'-half clones; Lane 5: clone with inappropriate insert; Lane 6: 1kb DNA ladder.

1.3. Cloning of viral hairpins

Because PiDnV hairpins within the viral ITRs cannot be sequenced by standard sequencing, we used a different strategy which is illustrated in Figure 12. Two clones of two halves of PiDnV genome (5'-half clone and 3'-half clone) and vector pBluescript KS (+) were used as starting materials to resolve hairpins at both ends of the viral genome.

1.3.1. Subcloning BamHI-BamHI fragment into pBluescript KS (+)

Two sites for BamHI were identified within the sequence of the one-half clones (5'-half clone or 3'-half clone), one in the multiple cloning site of the vector and the other at nucleotide 270 of the viral genome. Cutting the clones with BamHI (Invitrogen) gave one band of about 289 bp and another high molecular weight band (around 5.8 kb for 5'-half clone and around 5.4 kb for 3'-half clone). Therefore, the BamHI-to-BamHI fragment within a one-half clone (clone A in Figure 12) was subcloned (cloning (i)) into the vector pBluescript (2.9 kb) at the BamHI site, resulting in the clone (B). The purpose of this cloning strategy was to reduce the size of the insert and thus to facilitate subsequent cloning and sequencing of the hairpins. The

positive clones were confirmed by digesting with BamHI, which gave two bands, a 2.9 kb band and a 289 bp band as shown in Figure 13.

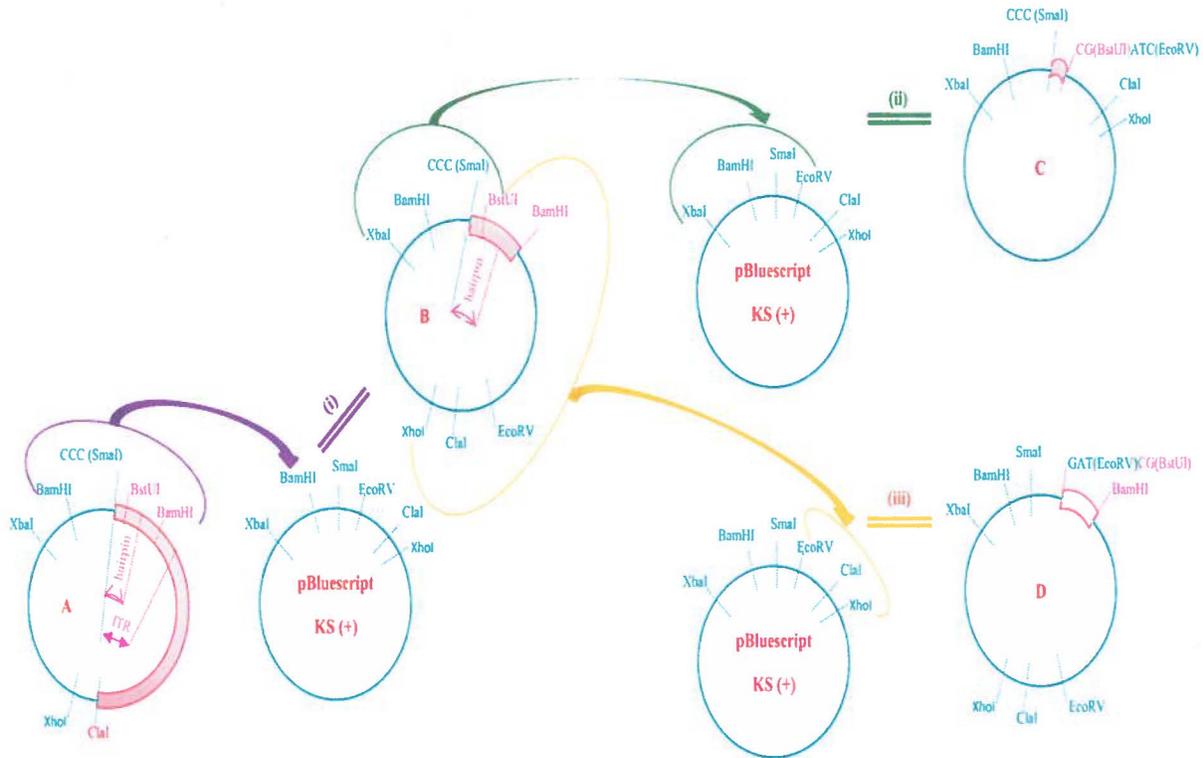


Figure 12. Cloning for sequencing of hairpins. In the diagrams and for both the sequence and the restriction endonuclease sites, blue is of vector, pink is of insert, red is of both vector and insert. The insert to be excised and cloned to another vector is indicated by an arc. This arc covers two restriction sites of the insert. An arrow connects the insert and the vector at the restriction site to be cloned. Distinct cloning is differentiated by 3 colors: purple, green and orange. A, B, C and D are the clones. (i), (ii) and (iii) indicate the clonings 1, 2 and 3.

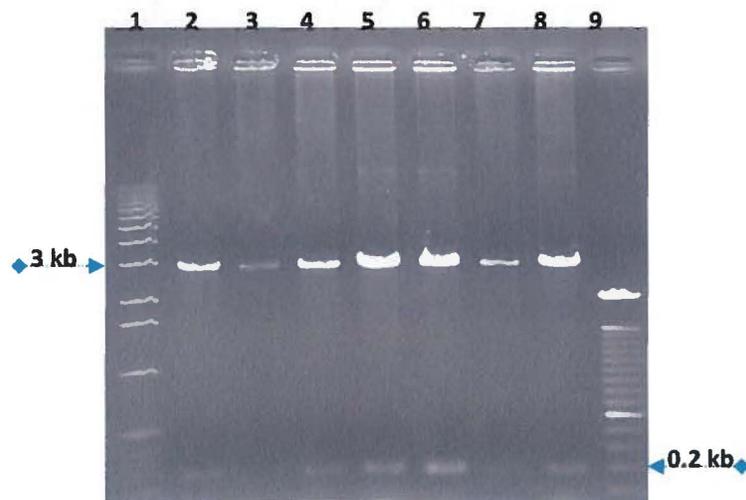


Figure 13. BamHI-BamHI subclone. Lane 1 is 1kb DNA ladder. Lane 2 to 8 are the clones digested with BamHI, positive clones gave two bands on gel: 2.9 kb and 289 bp bands. Lane 9 is 100 bp DNA ladder.

1.3.2. Cloning of two parts of hairpins

Digesting the 289-bp BamHI-to-BamHI fragment with BstUI (CG[^]CG, blunt enzyme, NEB) resulted in two bands which are approximately 85 bp and 204 bp, indicating that there is a BstUI site in the hairpin because it was not found in the BamHI-to-SmaI sequence of the vector. Therefore we used this enzyme to separate the hairpin halves within the BamHI-BamHI clone (clone (B)) in order to make it amenable for sequencing. In so doing, two fragments were cut out from the clone (B), the XbaI-to-BstUI and the BstUI-to-XhoI.

After digestion, two bands (260 bp and 92 bp) were obtained for 3'-half clone and two other bands (225 bp and 127 bp) for 5'-half clone (Figure 14). The four fragments were then cloned separately into pBluescript KS (+) digested with either XbaI-EcoRV or EcoRV-XhoI. The clones with proper inserts were identified by digestion with PvuII (5'-GAG[^]CTG-3') (NEB) which resulted in the band of 2.5 kb (vector) and one of the bands of 499 bp, 680 bp, 632 bp, 547 bp for one of the clones of fragments 92 bp, 260 bp, 225 bp, 127 bp, respectively (Figure 15).

As a result, we got the clone (C) for the XbaI-to-BstUI fragment (cloning (ii)) and clone (D) for the BstUI-to-XhoI fragment (cloning (iii)) in Figure 12). These clones were used for sequencing the viral hairpins.

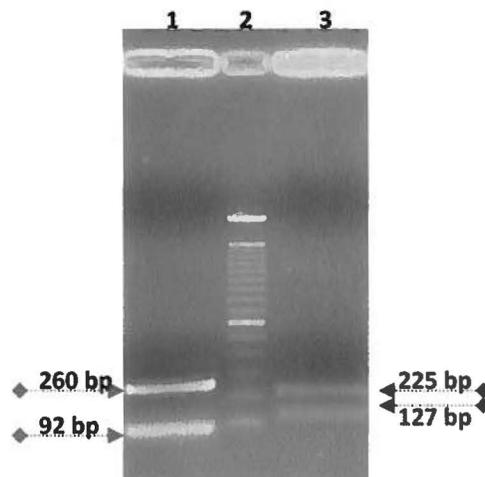


Figure 14. Digestion of BamHI-BamHI clones. Lane 1: 3'-clone. Lane 2: 100 bp DNA ladder. Lane 3: 5'-clone.

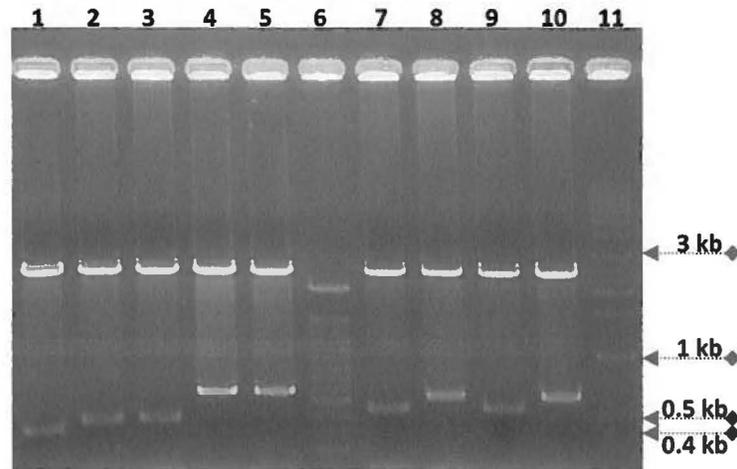


Figure 15. Hairpin clones (digested with PvuII). Lane 1: pBluescript KS (+). Lane 2 and 3: clones of 92 bp. Lane 4 and 5: clones of 260 bp. Lane 6: 100 bp DNA ladder. Lane 7 and 9: clones of 127 bp. Lane 8 and 10: clones of 225 bp. Lane 11: 1 kb DNA ladder.

1.4. Cloning the whole genome of PiDENV

The complete genome of PiDENV was cloned using the linear vector pJAZZ-OC (Lucigen Corp.). This vector is linear and is comprised of 2 arms (2 kb and 10 kb). It has also a chloramphenicol-resistant gene on one arm and the replication-related genes on the other arm which allows screening of recombinant clones.

To identify positive clones with the complete viral genome, at first 48 minipreps were prepared from 48 single colonies. One of these colonies seemed to have the insert which was then digested with NotI (NEB) which gave 2 arms of the vectors (2 kb and 10 kb) and the insert (6 kb). This digestion (Figure 17) confirmed the presence of the viral genomic DNA (6 kb) in the positive clone (lane 4).

To have more positive clones, 1800 single colonies were streaked on LB-Ampicillin plate. A sterile toothpick was used to make a cross across the streak of bacteria and the tip with bacteria was suspended in 10 μ l of 1 x Taq buffer (Invitrogen), ten single colonies were suspended in the same tube. 5 μ l of this suspension was used as template for colony PCR. In addition to conventional PCR, we used paraffin wax to improve specificity and sensitivity of PCR (32). 5 μ l of each tube of PCR products were loaded on agarose gel for visualizing the amplicons whose size would be 1 kb by electrophoresis. A 1-kb band was observed in 23

samples, i.e. 230 potential single colonies. Examples of colony PCR were illustrated in Figure 16.

These 230 single colonies were cultured and then a non-ionic detergent (NID) miniprep plasmid isolation (41) was performed. Putative positive clones were digested with NotI to confirm the presence of the appropriate insert. Only 5 positive clones were obtained of which the NotI digests gave 2 bands of the vector (2 kb and 10 kb) and the band of the viral genomic DNA (6 kb), identical to the first positive clone (lane 4, Figure 17).

As a consequence, we obtained only 6 positive linear PiDENV genome clones from 1848 single colonies (around 0.3%).

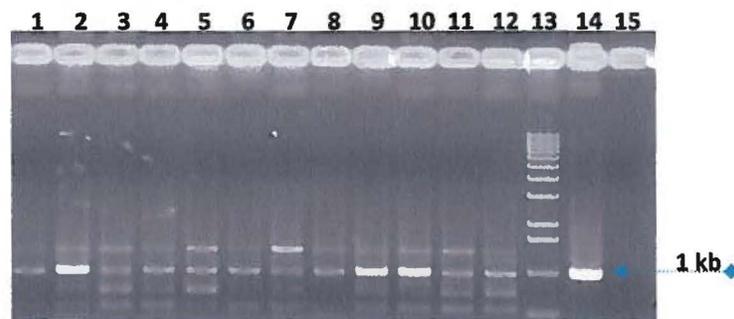


Figure 16. Example of colony PCR of the cloning of the whole genome of PiDENV. Lane 1 to 12: samples. Lane 13: 1 kb DNA ladder. Lane 14: positive control. Lane 15: negative control.

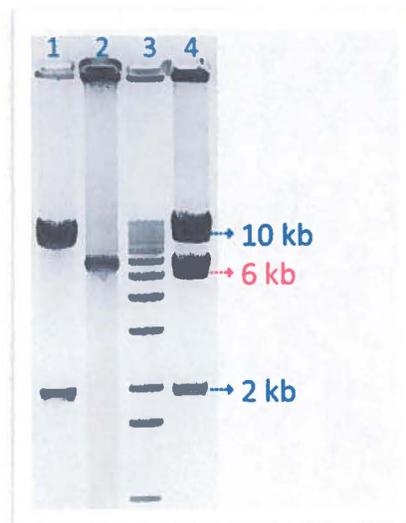


Figure 17. Positive linear clone of complete PiDENV genome (NotI). Lane 1: vector pJAZZ-OC. Lane 2: PiDENV genome. Lane 3: 1kb DNA ladder. Lane 4: the positive clone.

1.5. PCR to amplify the 5'-hairpin

Although a band of around 0.6 kb was observed for two clones (lanes 4 and 5, Figure 18), overall the first PCR of 5'-hairpin gave non-specific bands for five linear clones and no band for one clone. Therefore, 0.5 μ l of each reaction of this PCR was used as template for the second (nested) PCR reaction in order to obtain specific bands. As indicated in Figure 19, three out of five clones (lanes 2, 4 and 5) had the right band (200 bp). These three bands were then excised and purified to sequence the hairpins.

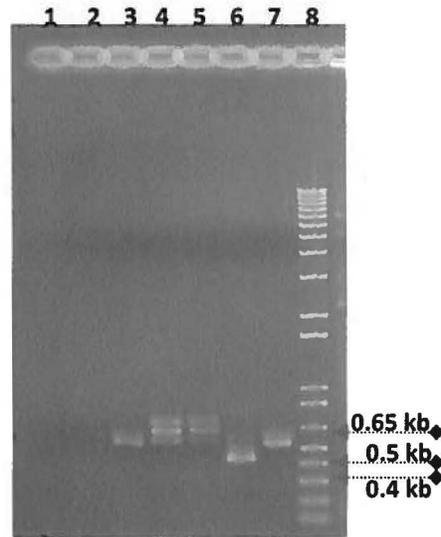


Figure 18. First PCR of the 5'-hairpin from linear clones. Lane 1: negative control. Lanes 2 to 7: linear PiDNV genome clones 1 to 6, respectively. Lane 8: 1 kb Plus DNA ladder (Invitrogen).

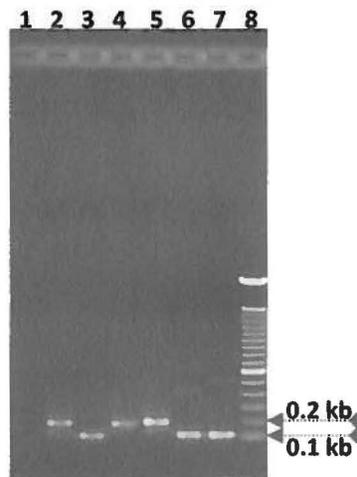


Figure 19. Second (nested) PCR of the 5'-hairpin from linear clones. Lane 1: negative control. Lanes 2 to 7: linear clones 1 to 6, respectively. Lane 8: 100 bp DNA ladder.

2. Sequence and organization of PiDNV genome

PiDNV genome was completely sequenced. The contigs were assembled by CAP3 program resulting in a sequence of 5990 nucleotides in length. When searching PiDNV sequence with BLAST (BLASTP 2.2.27+, NCBI) (3, 4), the highest identity was found between PiDNV and viruses in subgroup A of the *Densovirus* genus ((JcDNV (87%), MIDNV (86%) and GmDNV (84%)) while PiDNV sequence is very dissimilar to other parvoviruses even to other densoviruses in other groups of the genera (<25%) (Table 2).

Subfamily	Genus	Species	Genome length	Aligned score
<i>Densovirinae</i>	<i>Densovirus</i>	JcDNV	5906 bp	87
		MIDNV	6034 bp	86
		GmDNV	6039 bp	84
	<i>Pefudensovirus</i>	PfDNV	5454 bp	< 25
	<i>Iteravirus</i>	BmDNV1	5076 bp	< 25
	<i>Brevidensovirus</i>	AaeDNV	3978 bp	< 25
<i>Parvovirinae</i>	<i>Erythrovirus</i>	B19	5596 bp	< 25
	<i>Dependovirus</i>	AAV2	4679 bp	< 25

2.1. Telomeres of PiDNV genome

The genome of PiDNV is flanked by two large inverted terminal repeats (ITRs) each of which is 540 nucleotides in length. The sequencing of the majority of ITRs' sequence encountered no problem like for other standard sequences. However, when an internal primer, which is around 180 nucleotides downstream of the 5'-end of the genome, was used to sequence both the 5'-half and 3'-half clone in the conventional circular pBluescript, the furthestmost nucleotides at the extremity of each ITR could not be sequenced. Usually there was either a gap between the viral sequence and that of the vector or the sequencing stopped after about 10 nucleotides. The same phenomenon was observed when using the universal primer of vector. The sequencing of this part of PiDNV genome repeatedly failed. Nonetheless, separation of the sequence into two parts (BstUI) and then cloning and sequencing them separately worked well. Indeed the sequencing of these clones gave clear reads revealing a sequence of 120 nucleotides for only the 3'-end.

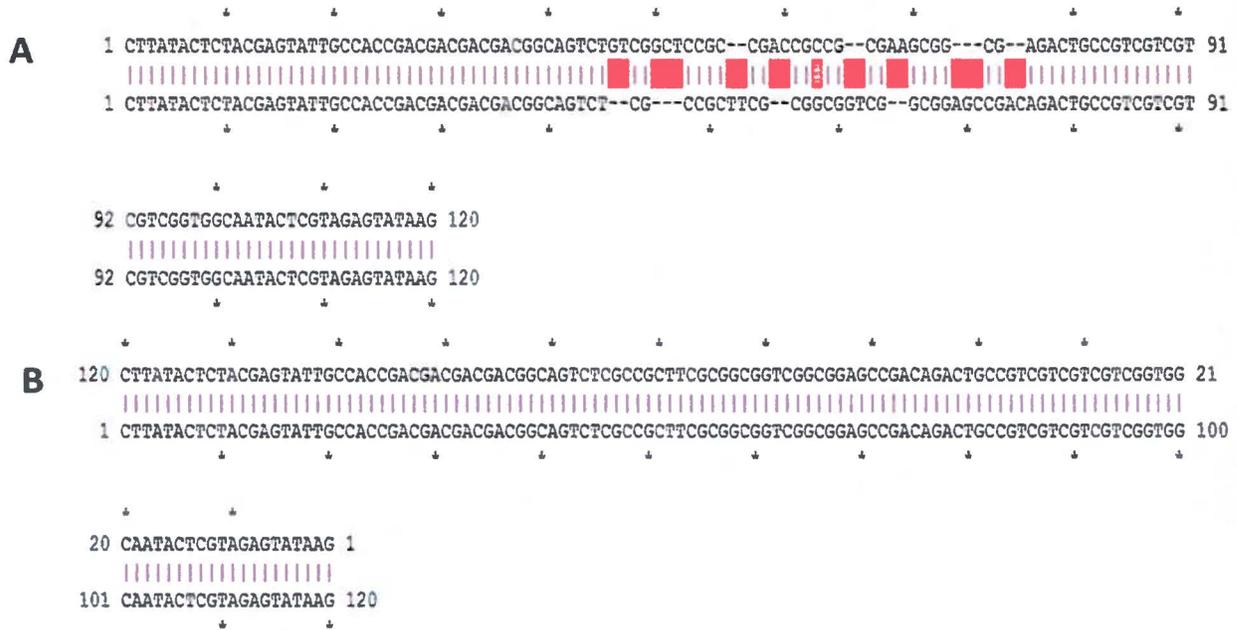


Figure 20. Alignment of two versions (flip and flop) of 3'-terminal hairpins. The upper sequence is the sequence of the 3'-hairpin of one clone and the bottom sequence is the sequence of the corresponding 3'-hairpin in a different clone. This resulted in either an imperfect alignment of flip and flop (with internal reverse-complement, between nts 45-76 in one hairpin, resulting in mismatching) (A) or a perfect alignment of flip/flip or flop/flop (B).

In order to have more data to confirm the sequence of 3'-hairpin as well as to have the sequence of the 5'-hairpin, linear pJAZZ cloning (section 1.4) and resequencing were done. As a result, the 3'-end of the viral genome was sequenced from 4 linear clones of which two clones had the same sequence as the first obtained sequence (circular clone). The sequences from two other clones possess only 101 nucleotides identical to the first sequence (Figure 20A). However, when reverse-complements of them were aligned with the first sequence, the identity percentage is 100% (Figure 20B). The first sequence (circular clone) is designated “flip” and the second “flop”. They are in fact reverse-complements. Therefore, we got 3 flip and 2 flop sequences for the 3'-hairpin. The structures of the flip and flop sequences of the hairpin were also observed using the mfold program for DNA (83) (Figure 21A). As shown in Figure 21, the hairpin of PiDENV had the Y-form, similar to those of MIDNV, GmDENV and JcDENV (66). In these hairpins, flip and flop sequences had identical stem whereas the sequences of the two “ears” are reverse-complements. In addition, the motif (GAC)₄ (22) were found at nucleotide 26 in both hairpins (Figure 21B) which can serve as a binding site for NS1 protein to create the origin of replication. For the 5'-hairpin, the standard sequencing of the linear clones did not

work. We have just recently solved it by using the method of Musso *et al.* (47) in which betaine, DMSO and 7-deaza dGTP were employed to facilitate the amplification of the hairpin (section 1.5). Consequently, the amplicons in three bands (200 bp) (Figure 19) from the three linear clones were excised and sequenced. Only 1 of them yielded good sequencing resulting in a sequence which is the reverse-complement of the 3'-hairpin sequence.

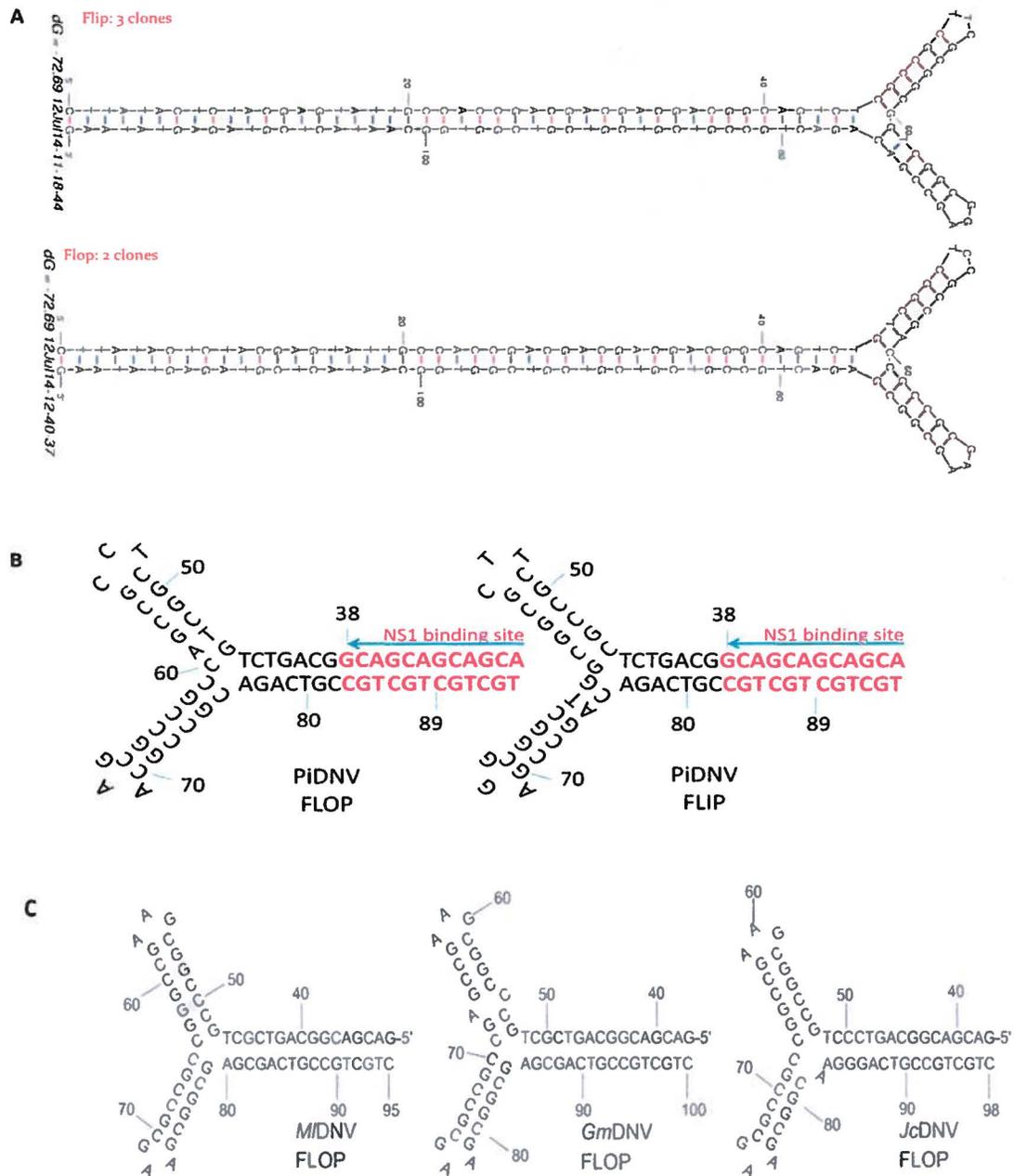


Figure 21. Hairpin of PiDENV. A. Secondary structure of PiDENV extremities predicted by mfold program (83). B. Redrawn hairpin of PiDENV with flop and flip configurations, NS1 binding sites are in red. C. Hairpins (flop) of MIDNV, GmDENV and JcDENV (66).

Within the ITRs, transcription elements were found (Table 3). Two promoters were predicted with the TATA box (consensus: TATATAA) located just at the inboard end of the ITRs. Besides, many binding sites for transcription factors were found in the ITRs. The names of these factors as well as the sites and nucleotide positions are listed in Table 3. Most of them are transcriptional activators such as AP-1, v-myb, TCF-1, etc.

Table 3. Predicted promoter sequence and transcription factors for these promoters (55, 60). In two promoter sequences, the TATA boxes are highlighted.

Promoter sequence	523-GACCATTGGGTATATAAAGTGATATTATTAGAGCGCCATCATCAGTCTAC-573 5466-CCATTGGGTATATAAAACAACCACGTGTTCTATTTTAGTCAGTATGTCCT-5416		
	Transcription factor	Binding site (motif)	Nucleotide position on PiDENV genome
5'→3'			3'→5'
CAP	CANYYY	501; 5695; 5701; 5749	290; 296; 487; 512; 543
CREB	TGACGTCA	157	5841
AP-1	GAGAGGA	265	5726
Lysozyme Silencer 1	ANCCTCTCY	272	5719
c-Myb	CMGTTR	5627	364
v-Myb	YAACKG	359	5632
Octa-U2snRNA	ATGCAAAT	426	5565
OTF-2A	TATGCAAAT	425	5566
H4TF-2	GGTCC	5682	526; 309
TCF-1	MAMAG	5625; 5762; 5785; 5816	366; 229; 206; 175
E2A	RCAGNTG	5626	365
GATA-1	MYWATCWY	437	5554
GATA-1	TATCTC	5593; 5614; 5642	398; 377; 349
GATA-1	TATCTT	439	5552
GATA-1	WGATAR	345; 373; 394; 5548	443; 5646; 5618; 5597
MTF-1	TGCRNC	5489	502
MBF-1	TGCRRC	5489	502
H-2RIIBP/T3R-alpha	GAGGTC	5538	458
W-box element	WGNAMCYK	304; 357; 499	5492; 5687; 5634
gamma-IRE	CWKKANNY	303; 413; 5738; 5769	514; 5826; 5801; 5688

2.2. Coding sequences of PiDENV genome

Four large open reading frames were found in PiDENV genome (ORF1-4) (Figure 22A). ORF1, ORF2 and ORF3 were encoded from 5' to 3' of the PiDENV sequence while ORF4 was encoded in reverse direction, i.e. 5' to 3' on the complementary strand of viral DNA. This organization is very similar to that of GmDENV, a highly related virus in the subgroup A of the

genus *Densovirus*. Like other viruses in the group, GmDnV has an ambisense genome organization (67) in which the non-structural proteins (NS1-3) were encoded from 5' to 3' and the structural protein (VP) was encoded by the largest ORF in the opposite direction. As presented in Figure 22, the distribution of ORF2, ORF3 and ORF1 of PiDnV was similar to those of NS1, NS2 and NS3 of GmDnV. The same can be seen for ORF4 of PiDnV and VP of GmDnV.

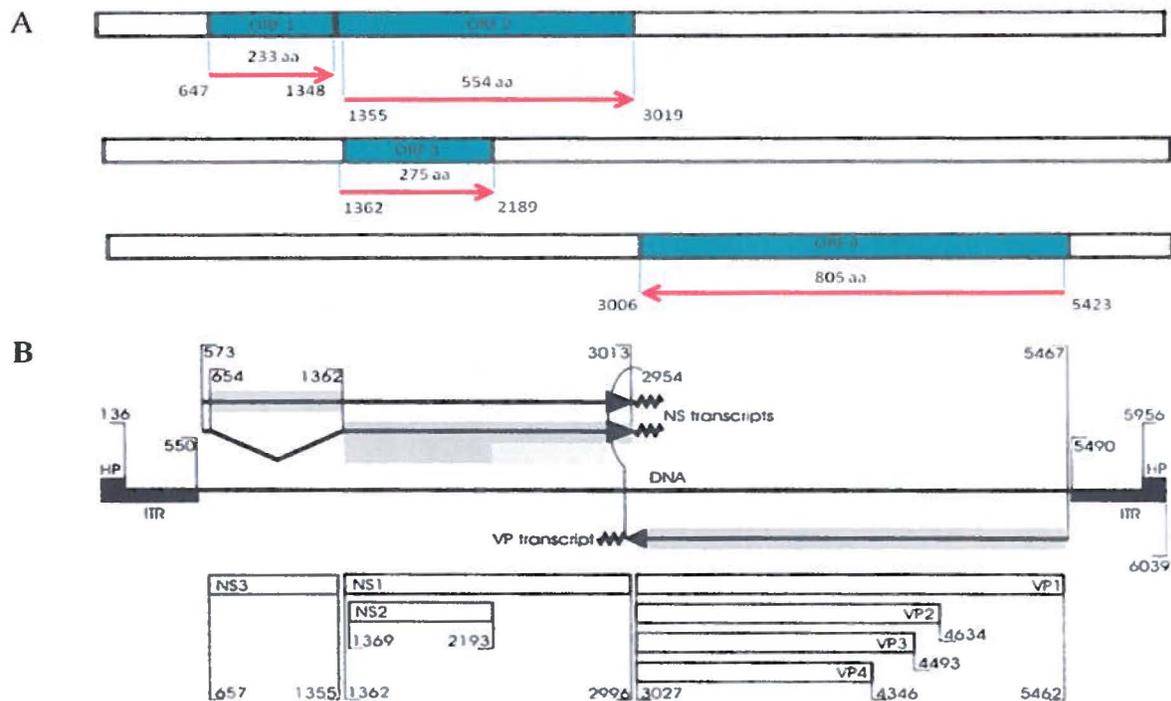


Figure 22. Genome organization. A. PiDnV genome organization. B. GmDnV genome organization (67)

2.2.1. Non-structural proteins

Non-structural proteins of PiDnV are potentially encoded by ORF1, ORF2 and ORF3 (Figure 22). The first ORF starts at nucleotide 647 and ends at nucleotide 1348 (702 nucleotides in length) which would give rise to a protein sequence of 233 amino acids (aa). When this sequence was blasted (BLASTP 2.2.27+, NCBI) (3, 4), high identity was found with NS3 protein of MIDNV (231 aa, 75%); GmDnV (232, 73%) and JcDnV (232 aa, 72%). Within this protein sequence, some important motifs were identified using ScanProsite (21) and Motif Scan (58) including zinc-finger motifs, N-glycosylation site, etc. (Table 4).

The second ORF begins at nucleotide 1355 and stops at nucleotide 3019 (1665 nucleotides, 554 aa). Within this ORF, a conserved domain (Conserved domain database-CDD) (44) for parvoviral non-structural NS1 protein was revealed (Figure 23). Especially the domain of the superfamily 3 helicase of DNA viruses was detected from amino acids 362 to 554 (Table 4). In addition, bipartite nuclear localization signal (amino acids 259 to 275) as well as N-glycosylation sites were also found in ORF2 (Table 4).

The third ORF is located within ORF2, from nucleotides 1362 to 2189 and codes for a sequence of 275 aa. Blast results (BLASTP 2.2.27+, NCBI) (3, 4) of this sequence showed a significant homology with non-structural NS2 protein of MIDNV (275 aa, 96%), JcDENV (275 aa, 93%) and GmDENV (274 aa, 94%).

2.2.2. Structural proteins

The remaining and largest ORF of PiDENV is ORF4. Starting only 27 nucleotides downstream of TATA box, the transcription of this ORF would be regulated by the transcriptional elements within the ITR. ORF4 is 2418 nucleotides in length (nucleotides 5423 to 3006) which encodes a sequence of 805 aa. Two conserved domains were found in this ORF (44). The first is of the densovirus capsid protein VP4 (pfam02336) (Figure 24) and the second is the N-terminal region of the parvovirus VP1 capsid protein (Figure 25). Within this conserved region of VP1, two important motifs were found which are PLA₂ (DxxAxxHDxxY) and Ca²⁺-binding loop (GPGN) (Figure 25). In addition, other conserved motifs were also identified using ScanProsite (21) and Motif Scan (58) (Table 4).

Table 4. Selected protein motifs and profiles present in ORF1, ORF2, ORF3 and ORF4.			
ORF	Predicted function	Motif	Amino acid position
1 (NS3)	N-glycosylation site	NRTV; NISV	163; 195
	cAMP- and cGMP-dependent protein kinase phosphorylation site	RKcS	208
	Zinc-finger	CkfCytnC ; CeiCHtC	145; 185
2 (NS1)	Superfamily 3 helicase of DNA viruses domain profile	NSLNIIIVELLKYqcnddedliveFLTNI VnvlrriPKLNAFLILSPSAGKNF FFDMIFGLLLSYGQLGQANRHNL FAFQEAPNKRVLWNEPNYESSL TD- TIKMMFGGDPYTVRVKNRMDA HVKRTPVIIITNNTVPFMYE YETAfSDRliqykwaapflkdyelkp hpmtffllskynisyfICVLFKHN----- ----	362
		Bipartite nuclear localization signal	KKAYSAGKFAYIRKTK
	N-glycosylation site	NRSC; NFTD; NNTV; NISF	145; 338; 496; 542
3 (NS2)	Bipartite nuclear localization signal	RKPIRQGNsHTYGKRQKR	257
	N-glycosylation site	NVSG; NQSS ; NLSE	41; 165; 172
	Amidation site	yGKR	268
4 (VP)	Ca ²⁺ -binding loop	GPGN	186
	Phospholipase A ₂	DxxAxxHDxxY	201
	N-glycosylation site	NQSE; NGSV; NTSD; NNSA	458; 467; 680; 790
	cAMP- and cGMP-dependent protein kinase phosphorylation site	KKfS	407
	Amidation site	fGKK	405
	Tyrosine kinase phosphorylation site	Keh.Dea.Y	205
	Protein kinase C phosphorylation site	TgK; SeR; SqK ; TkR; SkK; ShK; SnR; TnR; SyK ; TaK ; SdR	86; 104; 166; 327; 340; 415; 482; 495; 626; 647; 682
	Copper binding octapeptide repeat	-HKGPWGA	703
	U3 small nucleolar RNA C terminal	FGTIFDIY--SDIEKSQVLHK	686



Figure 23. Conserved domain within the ORF2. Sequence 16574 is of PiDNV and pfam01057 is of parvovirus non-structural protein NS1.

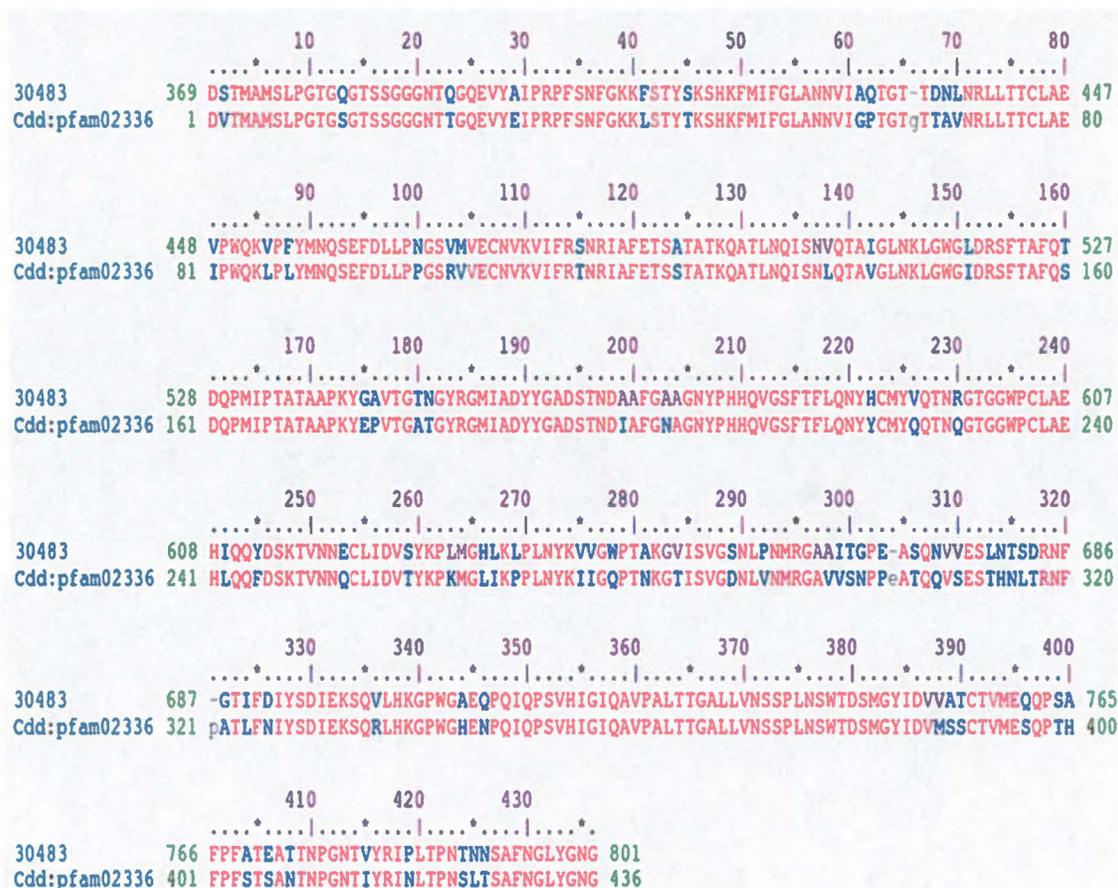


Figure 24. Alignment of the conserved domain within ORF4 (30483) with the conserved domain of capsid protein VP4 of densovirus (PFAM02336).



Figure 25. Alignment of phospholipase A₂ domain in VP1 of PiDNV and representative parvoviruses. Accession numbers of GmDENV, MIDNV, MVM, PPV, B19 are NP_694830, NP_958101, AAA671111, P18546, P07299, respectively. Conserved amino acids were shown in pink.

3. Predicted structure of capsid subunit of PiDNV

We used Swiss-Model with automated mode to generate the model of PiDNV capsid subunit. The reliability of this modelling is exemplified by the modelling of MVM and PPV of which structures were determined by X-Ray crystallography (MVM (PDB=1MVM;

Structure 15, 1369 (coll. Tattersall (Yale) and Rossmann (Purdue)) and PPV (PDB=1k3v; J. Mol. Biol. 315, 1189 (coll. Tijssen (INRS-IAF) and Rossmann (Purdue))). Although sequence identity of these two sequences is only 52% (Figure 26), superposition showed a high similarity between the two structures (Figure 2D). Therefore, we expect to have good prediction of PiDNV capsid subunit structure using GmDNV as template because the VP sequence identity of these viruses is up to 81%, much higher than that of MVM and PPV.

The 3D structure of VP of GmDNV was already resolved (59). Thus PiDNV VP sequence was aligned with that of GmDNV to get the corresponding sequence (aa 22 to 426) (Figure 27). This sequence was then submitted on the server of Swiss-Model with automated mode and template of GmDNV (1DNV, Chain A). The reliability of generated PiDNV model was estimated by QMEAN4 global scores. As shown in Figure 29, the model has a very low QMEAN4 raw score (0.308) and a strongly negative (-7.76) QMEAN Z-score. This means a low quality of the model. In addition, the reliability of this model was also estimated per residue by colour gradient. As depicted in Figure 30, blue colour was observed within the core of the model which indicates a high reliability. In contrast, the N-terminus, loop 2, loop 3 and other loops in the periphery are rather in red suggesting a low trustworthiness.

The homology modelling also gave the co-ordinates of atoms (Figure 28) which were used to generate the model with the UCSF Chimera software (50). This stand-alone model of PiDNV capsid subunit structure (Figure 31) is identical to that created by Swiss-Model (Figure 30) with the less reliable outside loops. However, when this model was integrated into the capsid (matrix for T=1 symmetry), external loops do fold and intertwine perfectly within the canyons of the capsid confirming a reliability of the loop structure and the capsid as a whole (Figure 32).

```

MVM  39  GVGVSTGSYDNQTHYRFLGDGWVEITALATRLVHLNMPKSENYCR
PPV  38  GVGVSTGTFNQTEFQYLGEGLVRITAHASRLIHLNMPEHETYKR
MVM  84  IRVHNTTDTSVKGNMAKDDAHEQIWTPWSLVDANAWGVWLQPSDW
PPV  83  IHVLNSE-SGVAGQMVQDDAHTQMVTPWSLIDANAWGVWFNPADW
MVM 129  QYICNTMSQLNLVSLDQEIFNVVLKTVTEQDSGGQAIKIYNNDLT
PPV 127  QLISNNMTEINLVSFEQEIFNVVLKTITESAT-SPPTKIYNNDLT

```

Figure 26. Alignment of amino acid sequence of VP of MVM and PPV. Homology is shown in red. Sequence identity is 52 %.


```

PIDNV 1 MSLPGTQGTS SGGGNTQGQEVYAIPRPFNFNGKKFSTYSKSHKFMIFGLANNVIAQTGT-TDNLNRLTTCLAEVPWQK 79
GmDNV 1 MSLPGTGS SGGGNTQGQDVYIIPRPFNFNGKLLSTYTKSHKFMIFGLANNVIGPTGTGTAVNRLTTCLAEIPWQK 80

PIDNV 80 VPFYMNQSEFDLLPNSVMVECNVYVIFRSNRIAFETSATATKQATLNQISNVQTAIGLNKLGWGLDRSFTAFTDQPMI 159
GmDNV 81 LPLYMNQSEFDLLPPGSRVVECNVYVIFRTNRIAFETSSTVTKQATLNQISNVQTAIGLNKLGWGINRAFTAFAQSDQPMI 160

PIDNV 160 PTATAAPKYGAVTGTNGYRGMADYYGADSTNDAAFGAAGNYPHHQVGSFTFLQNYHCMYVQTNRTGGGWPCLAEHI QQY 239
GmDNV 161 PTATAAPKYEPVTGDTGYRGMADYYGADSTNDAFGNAGNYPHHQVSSFTFLQNYCYMQTNQRTGGGWPCLAELQQF 240

PIDNV 240 DSKTVNNECLIDVSYKPLMGHLKPLNLYKVVGWPTAKGVISVGSNLPNMRGAAITG-PEASQNVVESLNTSDRNF GTI F 317
GmDNV 241 DSKTVNMQCLIDVTYKPKMGLIKSPLNYKIIIGQPTVKGTSVGDNLVNMARGAVVTNPPPEATQNVAEETHNLTRNFPADLF 320

PIDNV 318 DIYSDEKSQLVHKGKPGWAEQPQIQPSVHIGIQAVPALTTGALLVNSSPLNSWTDSDMGYIDVAVATCTVMEQQPSAFPAT 397
GmDNV 321 NIYSDEKSQLVHKGKPGWGHENPQIQPSVHIGIQAVPALTTGALLI NSSPLNSWTDSDMGYIDVMSSCTVMEAQPTHFPFST 400

PIDNV 398 EATNPGNTVYRILTPNTNNSAFNGLYGNG 426
GmDNV 401 EANTNPGNTI YRINLTPNSLTSAFNGLYGNG 431

```

Figure 27. Alignment of amino acid sequence of VP4 of PiDNV and GmDNV (COBALT) (48). Identical amino acids are in pink and identity is 81%.

ATOM	3094	N	GLY	A	404	17.075	-30.005	101.866	1.00	2.46	N
ATOM	3095	CA	GLY	A	404	16.489	-29.115	102.889	1.00	2.46	C
ATOM	3096	C	GLY	A	404	17.598	-28.876	103.922	1.00	2.46	C
ATOM	3097	O	GLY	A	404	18.691	-29.435	103.798	1.00	2.46	O
ATOM	3098	N	ASN	A	405	17.300	-28.120	104.964	1.00	2.37	N
ATOM	3099	CA	ASN	A	405	18.295	-27.963	106.032	1.00	2.37	C
ATOM	3100	C	ASN	A	405	17.884	-28.837	107.200	1.00	2.37	C
ATOM	3101	O	ASN	A	405	16.735	-28.850	107.609	1.00	2.37	O
ATOM	3102	CB	ASN	A	405	18.444	-26.509	106.468	1.00	2.37	C
ATOM	3103	CG	ASN	A	405	19.105	-25.666	105.388	1.00	2.37	C
ATOM	3104	OD1	ASN	A	405	20.311	-25.728	105.214	1.00	2.37	O
ATOM	3105	ND2	ASN	A	405	18.291	-24.923	104.651	1.00	2.37	N
ATOM	3106	N	GLY	A	406	18.844	-29.603	107.711	1.00	2.46	N
ATOM	3107	CA	GLY	A	406	18.572	-30.474	108.873	1.00	2.46	C
ATOM	3108	C	GLY	A	406	17.975	-29.592	109.984	1.00	2.46	C
ATOM	3109	O	GLY	A	406	18.234	-28.373	109.928	1.00	2.46	O
ATOM	3110	OXT	GLY	A	406	17.190	-30.107	110.791	1.00	2.46	O
TER	3111		GLY	A	406						O
END											

Figure 28. Example of co-ordinates of last atoms in the protein string of capsid subunit of PiDNV. Column 1: "Atom" with TER: Terminal atom; column 2: Atom serial number; column 3: Atom name; column 4: Residue name; column 5: Chain identifier; column 6: Residue sequence number; column 7: Orthogonal coordinates for x in angstrom; column 8: Orthogonal coordinates for y in angstrom; column 9: Orthogonal coordinates for z in angstrom; column 10: Occupancy; column 11: Temperature factor; column 12: Element symbol.

Scoring function term	Raw score	Z-score
C_beta interaction energy	-10.37	-2.41
All-atom pairwise energy	-665.08	-3.83
Solvation energy	15.31	-4.86
Torsion angle energy	9.43	-5.33
QMEAN4 score	0.308	-7.76

Figure 29. Reliability of PiDNV capsid subunit model: QMEAN4 global scores. QMEAN4 is a reliability score for the whole model which ranges between 0 and 1 with higher values for better models. QMEAN Z-score represents a measure of the absolute quality of a model by providing an estimate of the 'degree of nativeness' of the structural features observed in a model and by describing the likelihood that a given model is of comparable quality to experimental structures. Models of low quality are expected to have strongly negative QMEAN Z-scores (7).

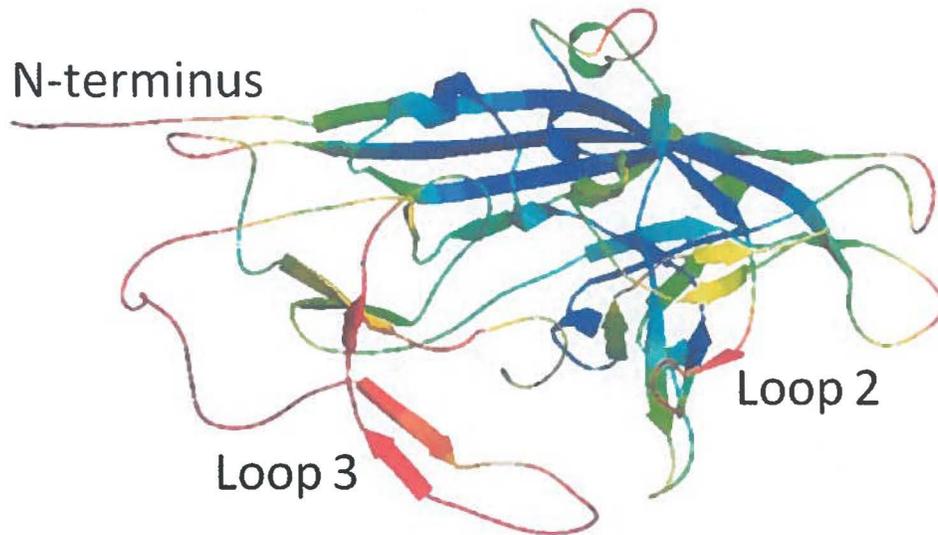


Figure 30. Reliability of PiDENV capsid subunit model according to Swiss Model (online): Colouring by residue error. Inaccuracy per residue is estimated by colour gradient from blue (expected to be reliable regions) to red (potentially unreliable region) (38).

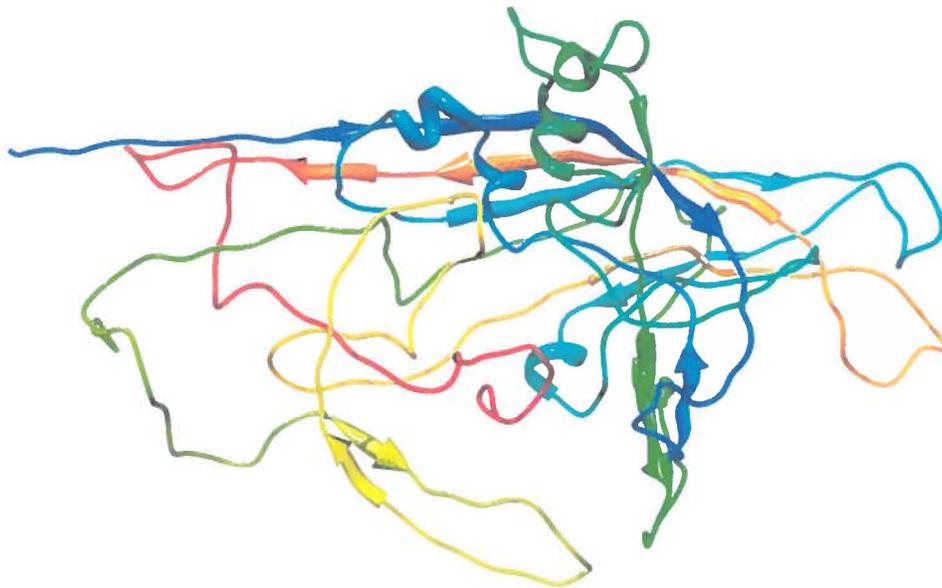


Figure 31. PiDENV capsid subunit. This model was built using the software UCSF Chimera (50) with the coordinates of PiDENV (Swiss-Model)

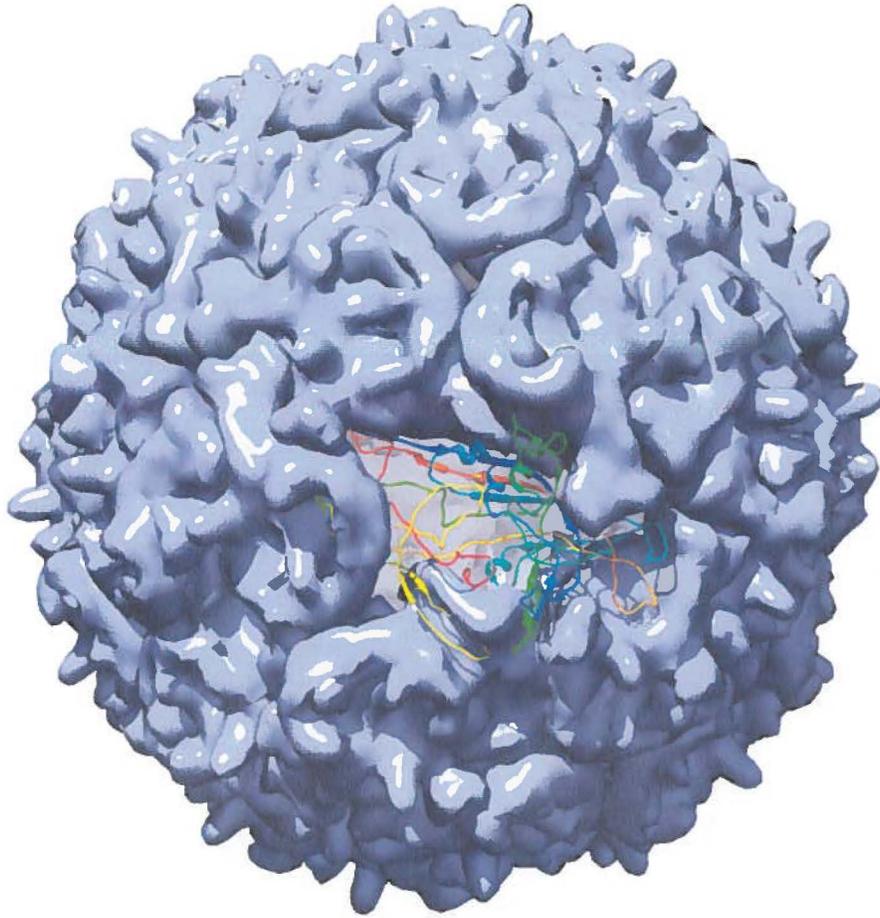


Figure 32. Model of capsid structure of PiDENV. The model was generated by UCSF Chimera using T=1 symmetry. Loops 2, 3 and the N-terminus fold precisely within canyons of capsid for optimal stability. The right-end of the inserted protein model is at the 5-fold axis.

DISCUSSION

1. Cloning and sequencing of PiD*NV* genome

With the conventional circular vector pBluescript KS (+), only halves of the viral genome can be cloned while the complete viral genome can be obtained using the linear vector pJAZZ-OC. The linear cloning is inefficient because we got only a very small number of good clones (about 0.3%). Indeed, the selection of positive clones in the linear cloning was very laborious. Nevertheless, the paraffin-colony PCR and NID minipreps were shown to be useful by reducing the workflow. The reason for the difficulty in both cloning is unknown at this time. It is possible that the repetitions of the extremities at two ends of the viral genomes as well as their high-GC content of the terminal hairpins make it difficult to clone the complete viral DNA.

It is well-known that cloning of parvovirus genome often yields clones with deletions or that have undergone recombination within the telomeres (5, 25, 28, 67). This is the same case for PiD*NV* telomeres. In fact, there were an additional 10 nucleotides to the 5' hairpin (the pBluescript clone) and a deletion of 13 nucleotides in one of the linear pJAZZ clone. Because four of the six clones in the linear pJAZZ vector had intact hairpins, it seems that this linear vector can decrease the deletion of viral sequences in recombinant plasmids inside the bacterial cells. The linear vector pJAZZ is known to be ideal for cloning of difficult sequences such as those with inverted repeats because of its two major characteristics. The first is that the vector is linear so that during replication, its two ends are free of torsional stress (which is the cause of instability of GC-rich sequences in conventional circular vectors). Secondly pJAZZ has two transcriptional terminators flanking the cloning sites which prevent transcriptional interference between the vector and insert and hence reduce further the instability and loss of inserts (31).

The sequencing of most of the DNA sequence of PiD*NV* was carried out without difficulty while that of the hairpins at two extremities of the genome encountered a lot of problems. The high-GC-content secondary structures like hairpins are well-known to be difficult to be sequenced because it can cause the DNA polymerase either to stop before intramolecular loops (53) or to jump over the loops (72). Therefore it is not surprising in the case of PiD*NV*'s hairpins which are composed of 61% GC. We indeed faced the same trouble while trying to sequence them. However, digestion with BstUI or using an internal viral primer to sequence the 3' hairpin within the linear clone did work well. Furthermore, the combined addition of betaine,

DMSO and 7-deaza-dGTP (47) succeeded in facilitating the amplification of the complete sequence of the 5' hairpin by altering its melting characteristics. More specifically, DMSO disrupts base pairing while betaine can act as isostabilizing agent that can equalize the contribution of AT and GC base pairing to the stability of the DNA duplex. In addition, because 7-deaza-dGTP makes one less chemical contact than dGTP, it can substitute dGTP in PCR and reduce the stability of the secondary structure (33, 36, 37, 47, 51, 54).

2. Nucleotide sequence of PiDENV genome and comparison with other viruses in the family

Telomeres of PiDENV are among the largest ITRs of the family *Parvoviridae*. This 540-nucleotide sequence displayed significant homology with ITRs of other ambisense densoviruses in the subgroup A (within the genus *Densovirus*) such as GmDENV (550 nucleotides) and MIDENV (543 nucleotides) with 89% and 93% identity, respectively. However there was no similarity between these sequences and those of parvoviruses in other genera. Similarly, the terminal hairpins are highly conserved between PiDENV and other subgroup A densoviruses (about 90%). Furthermore, although the hairpin sequences of these viruses are not completely identical, they shared the same Y-shape configuration. Finding flip and flop hairpins suggests strongly that PiDENV, like other parvoviruses (19), uses the rolling-hairpin replication mechanism to replicate its genome. Besides, NS1-binding sites within the hairpin stem are probably where the protein NS1 binds and then nicks the duplex DNA to create the origin of replication. Curiously, the NS and VP promoters including their TATA boxes are contained within the two ITRs as well as binding sites of transcription factors that potentially regulate the transcription of non-structural proteins and structural proteins of PiDENV. Temporal regulation of expression of these two gene cassettes would thus depend on sequences in the unique region.

3. Genome organization of PiDENV and comparison with other densoviruses

Similarity of sequence and genome organization of PiDENV and GmDENV suggests that PiDENV employs the same expression strategy as GmDENV. As revealed by the study of Tijssen *et al.* (67), non-structural proteins are encoded by three ORFs on the left hand of the genome. The first ORF codes for NS3 which is translated from a 2.5-kb unspliced NS transcript.

On the other hand, two other ORFs code for NS1 and NS2 from a spliced transcript of 1.8 kb by a leaky scanning mechanism. The three ORFs on the left hand of the PiDENV genome display the same organization. In addition to that, sequence similarity and conserved domain of superfamily 3 helicase on ORF2 suggest that ORF1 of PiDENV encodes NS3 and ORF2 and ORF3 encode NS1 and NS2. Like GmDENV, translation of NS1 and NS2 may be from a spliced transcript. 5'donor site (nucleotide 749) and 3'acceptor site (nucleotide 1354) can be employed to splice out the upstream NS3 to generate the NS1/2 transcript. The ORF of NS2 located within that of NS1 but their initiation codons (AUG) are at different places (nucleotide 1362 and 1355, respectively), thus translation could be by a leaky scanning mechanism. Moreover, the polyadenylation signal (AATAAA) for NS proteins is predicted to start at nucleotide 3002, 17 nucleotides upstream of the stop codon of NS1.

Non-structural proteins of PiDENV can have many roles in the viral life cycle. Like NS1 of JcDENV (22), NS1 of PiDENV can bind to the (GAC)₄ sequence in the terminal hairpins and nick the DNA during viral genome replication. Furthermore, a conserved helicase-domain in this protein suggests that PiDENV's NS1 also has helicase activity. Additionally, the nuclear localization signal (Table 4) can be employed to transport the protein to the cell nucleus to exert its functions. Intracytoplasmic proteins, such as NS proteins, are not prone to glycosylation. In contrast, the role of NS2 has not yet been identified although motifs like those in the overlapping part of NS1 sequence were also found (Table 4). NS3 of JcDENV was proved to play an essential role in replication of the virus (1) and it could be the same case for NS3 of PiDENV. In addition to other motifs (Table 4), a zinc-finger motif was also identified in NS3 of PiDENV at position 145 (CkfCytnC) and 185 (CeiCHtC).

The largest ORF4 of PiDENV potentially encodes viral structural proteins. Indeed, conserved domains for VP1 of parvoviruses and VP4 of densoviruses (Figure 24 and 25) were identified using CDD (NCBI) (44). Similar to group A ambisense densovirus (25, 67), the single transcript from this ORF can be translated into four structural proteins (VP1-4) by a leaky scanning mechanism. Within the VP1 sequence, the PLA₂ motif was found highly conserved (Figure 25). The Ca²⁺-binding loop (GPGN) is at amino acid 186 of VP1 of PiDENV, which is 17 amino acids upstream of the HD motif at amino acid 207. Additionally, the sequence PGYKYL before the GPGN is conserved for PiDENV, GmDENV, MIDENV, MVM, PPV whereas for B19 only P, G and Y were conserved (PGxxYx) (Figure 25).

4. Prediction of the 3D structure of the capsid subunit and capsid of PiDENV

The 3D structure of a protein is essential for knowing its biological function. For experimentally resolving protein structures, X-ray crystallography and Nuclear Magnetic Resonance (NMR) are usually used. However, these methods are costly and time-consuming; *in silico* protein structure prediction approaches have been extensively developed. Homology modelling is one of them. This technique is based on the alignment of the amino acid sequence of the target protein with that of a protein with known structure. It might be the most accurate method for building the 3D structure of a protein where experimentally determined structure of a related protein is available (38). Therefore, this method was chosen to predict the 3D structure of PiDENV capsid subunit based on the resolved structure of the related dengue virus, GmDENV (81% sequence identity).

The modelling produced a model with the β -barrel domain being highly reliable. This can be explained by the arrangement of the beta-strands in an antiparallel fashion to create a very tight structure. The β -strands are connected by long loops. In GmDENV, these loops create most of the contacts between capsid subunits (59). The predicted structure is a single subunit where there is no contact with neighbouring subunits. This is maybe the reason why the generated model was considered less reliable in these loops.

The visualization (UCSF Chimera) of the capsid subunit in the whole viral particle showed the interactions of the loops, and close fitting, with other subunits. Thanks to these contacts, the folding of the loops was deemed accurate and hence, the predicted structure is more stable and accurate.

CONCLUSION

In conclusion, the genome of PiDENV was completely sequenced. The viral nucleotide sequence and genome organization suggest the classification of PiDENV to the subgroup A ambisense densoviruses in the genus *Densovirus* of the subfamily *Densovirinae* in the family *Parvoviridae*.

The viral genome is a single-stranded linear DNA which is 5990 nucleotides in length. Blast results showed high similarity between PiDENV genome and those of subgroup A ambisense densoviruses (84-87%) but much lower identity to other parvoviruses. Flanking the genome are two large ITRs with 540 nucleotides each. The 120 nucleotides at the extremity of each ITR can fold back onto themselves to create a Y-shape hairpin. This configuration is also found in hairpins of other densoviruses in the subgroup A as well as AAV and B19. There are two versions of the viral hairpins, flip and flop, which are a reverse-complement of each other. It suggests that PiDENV employs the rolling-hairpin replication mechanism, common among parvoviruses, to replicate its genome. Binding sites for NS1 protein are also present in the hairpins indicating that replication origin is located within this sequence. Additionally, promoter elements and transcription factor binding sites were also identified in the ITRs, which can regulate transcription of non-structural and structural viral proteins.

PiDENV genome possesses an ambisense organization. The first three ORFs, which start from the left of the genome, encode non-structural proteins NS1, NS2 and NS3. NS3 may be translated from an unspliced transcript while NS1 and NS2 are possibly translated by a leaky scanning mechanism from a NS3-spliced transcript. Conserved domains of helicase were found on NS1 suggesting the same function for NS1 of parvoviruses. Although there is no dominant conserved domain in NS3, this protein may be crucial for viral replication like in the case of NS3 of JcDENV (1). On the other hand, structural proteins of PiDENV were coded in reverse direction on the complementary strand of the genome. Four different species of PiDENV structural proteins might be produced by leaky scanning from the single large transcript VP. Within this sequence there are two important conserved domains which are of densoviral VP4 and parvoviral VP1. The conserved VP1 domain contains the PLA₂ motif which is present in most parvoviruses. From the VP sequence, structure of PiDENV capsid subunit was predicted which showed a structure highly similar to that of the related densovirus GmDENV.

Appendix 1. Publication

The sequence of PiDENV genome was submitted to the GenBank (NCBI) with the accession number as JX645046. The work was also published by Oanh T. H. Huynh, Hanh T. Pham, Qian Yu and Peter Tijssen, *Journal of Virology*, volume 86, 2012 (35). This paper is included below. The contributions are as follow: Peter Tijssen, my supervisor; I (Oanh T. H. Huynh) did the experiments; Hanh T. Pham and Qian Yu were involved in a similar project on JcDENV that is soon to be published. We encountered similar problems and collaborated to solve them. I will be a co-author on their paper on JcDENV.

Pseudoplusia includens Densovirus Genome Organization and Expression Strategy

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

J. Virol. 2012, 86(23):13127. DOI: 10.1128/JVI.02462-12.

Updated information and services can be found at:

<http://jvi.asm.org/content/86/23/13127>

These include:

REFERENCES

This article cites 13 articles, 6 of which can be accessed free at:

<http://jvi.asm.org/content/86/23/13127#ref-list-1>

CONTENT ALERTS

Receive: RSS Feeds, eTOCs, free email alerts (when new articles cite this article), [more»](#)

Information about commercial reprint orders: <http://journals.asm.org/site/misc/reprints.xhtml>

To subscribe to to another ASM Journal go to: <http://journals.asm.org/site/subscriptions/>

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 picomavirus 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 (MIDNV) (4). The large ORF1 (nt 1355 to 3019) on the plus strand had a sponded to NS2 (nt 1362 to 3019) with 275 aa, and ORF3 (nt 647 coding capacity for NS1 of 554 amino acids (aa), ORF2 corre-

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](https://www.ncbi.nlm.nih.gov/nuclot/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 loopure-baited traps. *J. Econ. Entomol.* **75**:647–650.
2. Chao Y-C, Young SY III, Kim KS, Scott HA. 1985. A newly isolated denonucleosis 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 denonucleosis virus from *Mythimna loreyi*. *Virology* **320**:181–189.
5. Godisla 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.

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

[doi:10.1128/JM102462-12](https://doi.org/10.1128/JM102462-12)

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 densovirus, 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 densovirus 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.

Appendix 2. Virus inactivation by Impulsive Stimulated Raman Scattering (ISRS)

In addition to the project “*Pseudoplusia includens* densovirus (PiDNV) genome organization and expression strategy”, I also carried out another project in my M.Sc. program. This project is “Virus inactivation using Impulse Stimulated Raman Scattering”. These two projects are completely different. Therefore I present the second project in Appendix 2 which comprises the literature review, methodology, results, discussion and conclusion.

This project is a co-operation between our laboratory and that at EMT. The experiments of the second project, which I present in the appendix 2, were done by me and Aziz Berchtikou (Master student at EMT). Aziz Berchtikou carried out laser experiments at EMT while I did the experiments on viruses at IAF, including the preparation of virus stock and virus infectivity test.

REVIEW OF LITERATURE

1. Inactivation technologies

Biological products have been playing important roles in worldwide healthcare systems. For example, blood and blood components that include red cells, plasma, platelets, albumin, immunoglobulin and coagulation factors have been employed for centuries in transfusion, as well as in the treatment of life-threatening diseases, such as anaemia, hemorrhage and haemophilia. However, clinical uses of these products are of high risk because they can contain a variety of blood-borne pathogens. According to the Centers for Disease Control and Prevention (CDC) (http://www.cdc.gov/bloodsafety/bbp/diseases_organisms.html), potential transfusion-transmitted organisms encompass a wide range of viruses, bacteria, parasites and prions. Strategies have been developed in an attempt to reduce these pathogens in the blood supply. A preliminary approach is the combination of donor selection, donor screening by questionnaire on risky behaviours and tests for known pathogens, such as Hepatitis B virus (HBV), Hepatitis C virus (HCV), Human immunodeficiency virus (HIV), Human T-lymphotropic virus (HTLV), and West Nile virus (WNV). As a result, reduction in the risk of infections by blood-associated pathogens was achieved such as the decreased prevalence of transfusion-transmitted hepatitis virus and HIV infections (23). Although this approach is effective, it is considered to be reactive. For example, many pathogens for which there are currently no screening tests available could potentially cause lethal infections. These include pathogens from both human and animals (zoonotics), dengue virus, SARS coronavirus, Deltavirus hepatitis D, human parvovirus B19, SV40, *Leishmania donovani* and prions, to name a few. Further, the emergence and re-emergence of infectious agents that have acquired new virulence factors, drug resistance or exhibit long latent periods in asymptomatic donors, have been posing potential threats to blood safety. In addition, it can take years from the discovery of the causative agent to implementation of a new preventative strategy and during this delay, infections might spread (2). For instance, after evaluating Factor VIII products from porcine blood for haemophiliacs, produced by a British pharmaceutical company, for several years our laboratory discovered the presence of a new pathogen PARV4 (64).

In contrast, pathogen inactivation is a proactive approach by which broad antimicrobial activity can prevent infections of both known and unknown pathogens. Over the last decades, many inactivation technologies have been developed some of which, such as

solvent-detergent (S-D) treated plasma, have already been applied in North America and Europe (9). Examples of these technologies are listed in Table 5. Despite the fact that implementing these technologies can reinforce the existing paradigm, limitations with regards to effectiveness, safety and cost can be a great hindrance for their application. In fact, currently available inactivation methods can target only a narrow spectrum of pathogens, one method can only reduce infections of a limited range of infectious agents. For example, S-D treatment is capable of inactivating only enveloped viruses, but fails to inactivate non-enveloped ones. Usually, one requires a combination of many methods. For example, Cohn fractionation, heat, pasteurization, S-D treatment and nanofiltration are used to produce coagulation factors and immunoglobulins (11). A single technology could require many complicated steps, expensive equipment and materials, so that using multiple technologies could increase the cost of the final products, making them unaffordable for many patients especially those in developing countries. Moreover, the introduction of exotic compounds to biological products has been raising a number of safety concerns. In many inactivation methods, substances used can bind unselectively to nucleic acids and even proteins. Therefore, not only are the pathogens targeted, but also our own nucleic acids and proteins. Consequently, they can cause damage to cellular nucleic acids and proteins, rendering therapeutic products ineffective or leading to mutation and cancer, once introduced to a patient's body. In addition, they can even confer resistance to infectious microbes by mutation, which can enhance their survival and pathogenicity (9). Furthermore, compounds like S-303 and PEN110 were reported to provoke specific antibody formation (6). Hence, these foreign agents would obviously induce immunity against beneficial therapeutic products. Finally, if the removal of residues and by-products after inactivation is not complete, adverse reactions could be generated following treatment. Due to these drawbacks, there is a strong and urgent need for better inactivation technologies.

2. Virus inactivation by ISRS

Recently, a new photonic approach was developed to inactivate viruses (71). This method used an ultrashort pulse fiber laser as the excitation source. The fiber laser emitted light at a wavelength of 1.55 μm , which was operated at a repetition rate of 500 kHz with an output of 5 μJ per laser pulse. The second harmonic of the 1.55 μm laser was used to inactivate bacteriophage M13, Human papilloma virus (HPV), Human immunodeficiency virus (HIV) and

Tobacco mosaic virus (TMV). The second harmonic is in the near-infrared (near-IR), with a wavelength of 776 nm, about 1.4 μJ per laser pulse, pulse width of about 600 femtosecond (fs) at full-width of half-maximum and a spectral width of about 70 cm^{-1} . While water strongly absorbs radiation at 1.55 μm (infrared) and thus the excitation laser energy cannot be transferred to the vibration energy of the viruses, it is highly transparent in the near-IR to the visible range. Irradiating the virus with femtosecond 776 nm laser was indeed successful in inactivating the viruses after 2 hours of irradiation at 25°C , with laser intensity of $100 \pm 10\text{ MW cm}^{-2}$. The physical mechanism of virus inactivation proposed by the authors is ISRS (Impulsive Stimulated Raman Scattering).

By *stimulated scattering*, the electric field of a laser light can act as a force that drives a *Raman*-active vibrational mode into motion in a particular direction, or in other words, excitation can occur. To have significant excitation, the duration of the laser pulses has to be short as compared to the time of a single vibration's oscillation. The lasers with such short pulses, such as femtosecond lasers, produce an *impulsive* driving force which can excite coherent acoustic photons in liquids and solids with Raman-active vibrational mode. The excitation is achieved by optical mixing of the Fourier components of the ultrashort incoming pulses (80). ISRS was reported to produce large amplitude of vibration modes in molecules within liquid solution as well as in solid systems (80). Like any mechanical object, viruses, and especially their capsids, have their own vibration modes (Raman-active vibrational modes), which are in the range of 30 GHz to 300 GHz (27, 69). Femtosecond laser pulses can therefore be used to excite the vibration of capsids of viruses through ISRS. The first laser pulse can bring the particles to a higher vibrational state. When subsequent pulses interact with the excited virus particles, the amplitude of the vibrations increases, achieving such a level that weak links between proteins within the viral capsid are broken down. As a consequence, capsid damage leads to virus inactivation (71).

Interestingly, the efficiency of this method is not affected by the genetic mutation of viruses. This is because the process targets vibration modes, or in other words, the global mechanical properties of viral particles (70). The approach is therefore very promising in confronting rapidly mutating viruses like HIV. Of equal interest is its selectivity. ISRS was shown to inactivate only viruses, while leaving other components, such as human red blood

cells, human Jurkat T cells and mouse dendritic cells, unharmed (71). In addition to the above-mentioned benefits, using a laser system is safe for the recipients and is cost effective. Since there is no exotic compound introduced to the samples, there is little chance that side effects could occur. Similarly, no immune response can be provoked due to neoantigens. Such an inactivation method can also be less complicated, since it can employ a universal apparatus and common protocol with adjustable parameters for a specific range of target pathogens. As a result, a scaled-up inactivation system can be cost-efficient.

Because of the multiple advantages of this inactivation method, our research also uses a femtosecond laser to inactivate viruses. However, the laser used in this work has much higher energy per pulse with a maximum peak power of 4 TW for the 10 Hz laser, 500 000 times more powerful than the laser used by Tsen *et al.* (71). In practice, this powerful laser can treat samples with large volume in a relatively short time, which can facilitate the scale-up of the process and reduce the cost for treatment. As there is currently no research using similar lasers, we have carried out experiments to determine important factors in the inactivation process. In this thesis, I will report on the preliminary experiments that I participated in.

Table 5. Representative inactivation technologies (9).

Technology	Principle of inactivation	Effectiveness
Cohn fractionation Heat treatment (14, 56)	Pathogen reduction after different steps of precipitation and physical separation (centrifugation, filtration). Heat treatment (100°C/1h).	Reduce HBV, HCV, HIV, Parvoviruses (56). Safe in last fractions (albumin, globulin). Earlier fractions (antihemophilic factors) need further treatment (heat, pH).
Solvent-Detergent (S-D) treatment (40)	Organic solvents and detergents disrupt lipid envelope of viruses, preventing from binding to cells and replicating.	Reduce enveloped viruses. Fail to inactivate non-enveloped viruses. Plasma protein concentrate and fresh frozen plasma made from S-D treated plasma. Decreased levels of coagulation factors.
Nanofiltration (10)	Size exclusion to remove viruses	4-6 log removal of most viruses. 90-95% recovery of protein activity.
Photoactive phenothiazine dye (Methylene Blue (MB), Thionine) (73, 74)	MB has high affinity to nucleic acids and surface of viruses. MB-treated plasma is exposed to UV light.	Inactivate most enveloped viruses. Non-enveloped viruses, intracellular viruses, bacteria, protozoa: not inactivated. Plasma proteins: moderately affected. Fibrinogen and factor VIII activity: 30% reduced (73, 78).
Psoralens (S-59) (43)	Intercalate between the bases of nucleic acids. After illuminated with UV-A, cross-linking is formed, prevent replication and transcription.	Inactivate viruses, bacteria, protozoa. Extensively tested in platelets and plasma. Acceptable safety profile. Possible platelet injury in a trial (46). Most coagulation factors: well preserved (43, 45).
Riboflavin/UV light (20)	Intercalate between the bases of nucleic acids. UV- or visible light exposure breaks riboflavin-linked nucleic acids.	Reduction in platelets: HIV, WNV, parvovirus B19, bacteria, protozoa. May increase glycolysis. Acceptable levels of clotting factors.
S-303 (Frangible anchor linker effectors compounds) (15)	Insert into helical region of nucleic acid. Shift from low to higher neutral pH caused hydrolysis	Used to treatment of red blood cells. Also bind to other proteins and cell membranes, 20% remained bound. Inactivate wide ranges of viruses, bacteria, protozoa (16). Antibodies to red cell bound S-303 (6).
PEN110 (inactine)	Bind to nucleic acid. Activation breaks nucleic acid strands (52).	Inactivate a broad range of viruses, bacteria, protozoa, mycoplasma in red blood cells. Red cell quality is maintained. Antibodies were found in a trial.
Leukoreduction filter	Filter.	Remove cell associated-viruses, prions (61).

METHODOLOGY

1. Preparation of virus stocks

We used either bacteriophages M13 or MS2 in the inactivation experiments. M13 is a filamentous virus that belongs to the family *Inoviridae* (68). It is a non-enveloped virus with a circular single-stranded DNA virus, with a size of about 900 nm long and 6-7 nm in diameter (Figure 33A). These phage infect *Escherichia coli* (*E. coli*) and propagate within their host as non-lytic viruses. Lentivirus MS2 is a non-enveloped lytic bacteriophage that also infects *E. coli* (Figure 33B). It is an icosahedral, positive-strand, single-stranded RNA virus which is only 27 nm in diameter (62). These two species of viruses were used because they do not infect humans and are easy to maintain in the laboratory. Their differences in size as well as shape can also be used to evaluate the impact of laser parameters on different kinds of viruses.

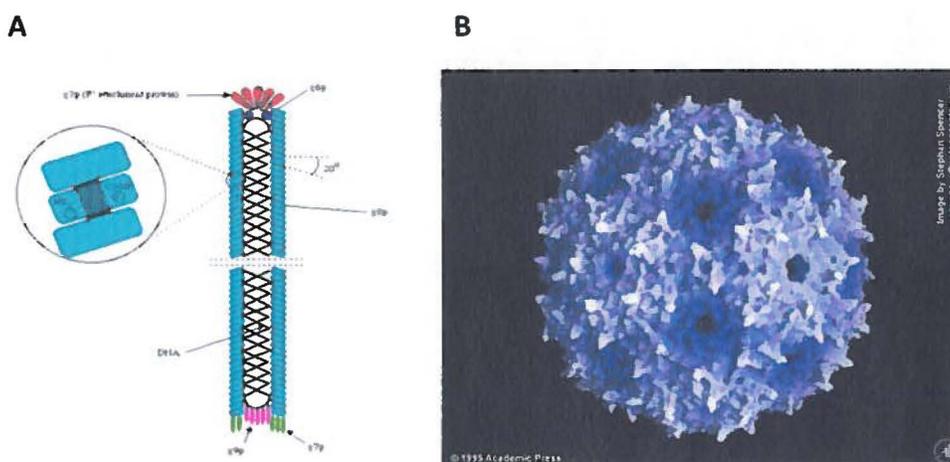


Figure 33. Viruses in inactivation experiments. A. M13 bacteriophage (<http://ntmf.mf.wau.nl/cor/m13.htm>). B. MS2 bacteriophage (<http://www.virology.wisc.edu/virusworld/imgency/ms2MS22.jpeg>).

To prepare virus stocks, a single plaque of phage was suspended in 1 ml of LB broth in a microcentrifuge tube. After gentle vortexing, the tube with the virus suspension was let stand at room temperature for 2 hours. 100 μ l of the virus suspension was then used to incubate 1.25 ml of host bacteria (strain JM103 for M13 and C3000 for MS2) in 50 ml of LB. The culture was incubated at 37°C with shaking at 250 rpm for 8 hours. The infected culture was centrifuged at 1200 x g for 10 minutes. After that, the supernatant including viruses was transferred to a glass bottle and stored at 4°C.

2. Experimental setup

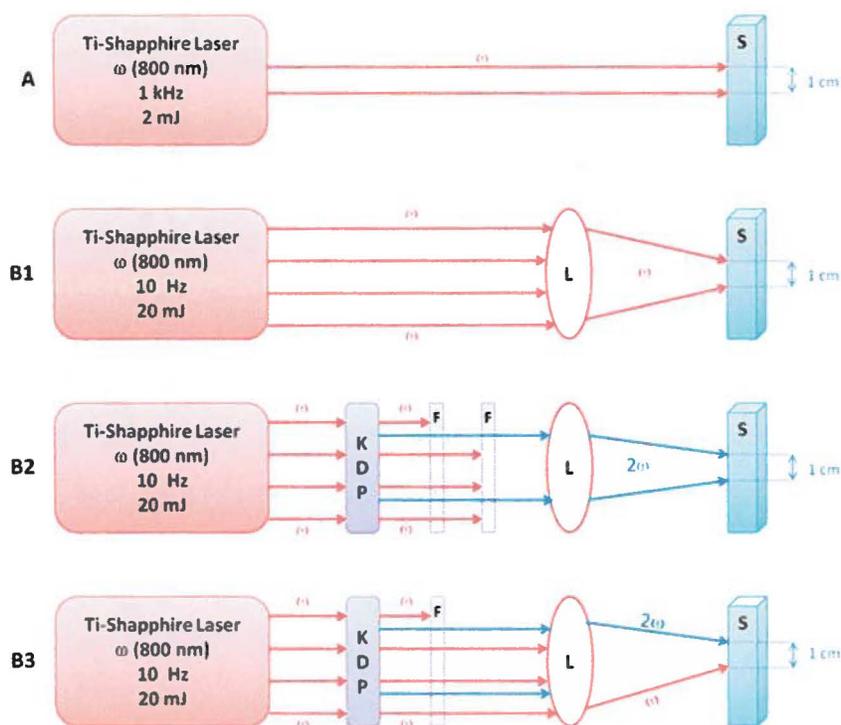


Figure 34. Experimental setup. Diagram A shows the schematic diagram of the experiment with the 1 kHz laser (2 mJ). Diagrams B1, B2 and B3 are for experiments with the 10 Hz laser at different wavelengths: B1: 800 nm (ω); B2: 400 nm (2ω). B3: 800 nm and 400 nm. KDP is the second harmonic generator, F is the filter (dichroic mirror), L is the focus lens ($f = 50$ cm) and S is the sample. The Blue arrow represents the laser at 400 nm and the Red arrow the laser at 800 nm.

The experimental setup is illustrated in Figure 34. The Ti:sapphire laser (800 nm) was used as the excitation source in all experiments.

First, we used the 1 kHz laser, with 35 femtosecond pulses, peak power of 57 GW and maximum energy of 2 mJ per laser pulse with a power density of 72.58 GW cm^{-2} , to irradiate directly (without focus) the glass cuvette (Labshops, pathlength = 10 mm) that contains either M13 or MS2 (Figure 34A). To evaluate the impact of the laser energy on virus inactivation, we used lasers with energies of 2 mJ, 1.5 mJ, 1 mJ and 0.5 mJ per pulse to treat viruses, in either 12 minutes or 24 minutes. To evaluate the impact of time, the samples were treated with the 1 kHz laser at maximum energy (2 mJ) for either 1 minute, 4 minutes, 10 minutes, 22 minutes or 46 minutes.

In a different experiment, we used the 10 Hz laser, with a maximum energy of 20 mJ per laser pulse (Figure 34B1-3). This time, all experiments were performed by focussing with

a long focal length lens ($f = 50$ cm) because the absence of focussing in the first experiment may be the cause of failure in virus inactivation. The viruses used were M13 (latter, MS2 were also tested under the same conditions, but since I was not a participant to that experiment, I will only report results with M13). These experiments were performed using laser output with two different wavelengths for the virus inactivation process. The first was the beam at 800 nm, which was focused on to the sample using the focus lens (Figure 34B1). The second beam was at 400 nm, which was generated by using a Potassium Dideuterium Phosphate (KDP) crystal and two dichroic mirrors (filters) (Figure 34B2). KDP is a second harmonic generator in which part of the 800 nm laser beam was converted into a 400 nm laser beam. Therefore, after the KDP crystal, two beams resulted from the process, which were then filtered by two dichroic mirrors. These mirrors function as filters that pass selectively a small range of wavelengths, yet reflect others. As a result, after the first dichroic mirror, some of the 800 nm light was reflected but all of the 400 nm passed. After the second dichroic mirror, all of the 800 nm was reflected and hence left only the 400 nm light, which then irradiated the samples. We also used the laser at two wavelengths, the 800 nm and the 400 nm (after the KDP), which was allowed to pass through a dichroic mirror before reaching the focus lens. The dichroic mirror in this setup splits the beam into two wavelength lasers with equal energy (10 mJ). The diameter of the laser beam that interacts with the samples is 1 cm in all experiments (Figure 34).

For the 800 nm or the 400 nm laser, we evaluated the impact of changing the irradiation time, energy and laser pulse duration on the inactivation efficiency. In experiments with both 800 nm and 400 nm, only the impact of different laser pulse durations was evaluated, while energy (20 mJ in total) and time (24 minutes or 60 minutes) were kept constant. The details of the values used are listed in Table 6. Peak power and Power density were calculated by the equations below:

$$\text{Peak power} = \frac{\text{Energy}}{\text{Pulse width}}$$

$$\text{Power density} = \frac{\text{Peak power}}{\text{Area of the laser beam}}$$

$$\text{Area of the laser beam} = \pi r^2$$

Here r is the radius of the laser beam and is equal to 0.5 cm (since the laser beam that interacts with the sample has a diameter of 1 cm).

The viruses in all inactivation experiments were contained in a glass cuvette (Labshops, pathlength = 10 mm) with a concentration of 10^8 pfu/ml and a volume of 1 ml/cuvette. A magnetic stirring bar (Fisherbrand) was used to promote the interaction of the laser photons with the virus particles only in experiments using the 10 Hz laser. All experiments were done at 25°C and the temperature of samples in a cuvette was measured right before and after each experiment, with a thermocouple (OMECA). Before or after the treatment, the samples were always stored at 4°C. Before each treatment, 200 µl of viruses were set aside to serve as untreated control.

Table 6. Experiments with the 10 Hz laser.

Wavelength (nm)		Time (minute)	Energy (mJ)	Pulse width (femtosecond)	Peak power (TW)	Power density (TW/cm ²)
800 or 400	Impact of inactivation time	12	20	40	0.5	0.6
		24				
		60				
	Impact of inactivation energy	24 or 60	20	40	0.5	0.6
			15		0.375	0.5
			10		0.25	0.3
			5		0.125	0.2
800	Impact of inactivation pulse width	24 or 60	20	40	0.5	0.64
				100	0.2	0.25
				250	0.08	0.10
				500	0.04	0.05
				750	0.03	0.04
				1000	0.02	0.03
				1250	0.016	0.02
				1500	0.013	0.017
400	Impact of inactivation pulse width	24 or 60	20	40	0.5	0.64
			15	100	0.15	0.19
			12	250	0.048	0.061
			8.3	500	0.0166	0.021
			6.3	750	0.0084	0.011
			5.6	1000	0.0056	0.007
			3.2	1250	0.0025	0.003
			2.6	1500	0.0017	0.002
800 and 400	Impact of inactivation pulse width	24 or 60	20	40	0.5	0.64
				100	0.2	0.25
				250	0.08	0.10
				500	0.04	0.05
				750	0.03	0.04
				1000	0.02	0.03
				1250	0.016	0.02
				1500	0.013	0.017

3. Plaque assays to determine virus infectivity

Because both M13 and MS2 are able to form visible plaques while growing on the host bacterial lawn, the infectivity of viruses before or after inactivation was determined by plaque assays. Untreated controls were tested along with laser-irradiated samples. Viruses were diluted 10-fold with sterile H₂O (15 µl of viruses and 135 µl of sterile H₂O) up until the dilution where plaques were well separated on an LB agar plate (100 x 15mm, Fisher Scientific).

After the dilution, 100 µl of each diluted virus sample or control was mixed with 100 µl of host bacteria in 3ml of melted LB Top Agar at 50°C. The mixture was then poured onto a LB Agar plate which is subsequently allowed to stand at room temperature for 10 minutes or until the agar solidified. The plates were then incubated at 37°C overnight. After incubation, the number of plaques on each agar plate was counted in plaque forming units (pfu). An example of a titration plate is illustrated in Figure 35. Each sample or control is tested three times, from which the average number was calculated and used for subsequent calculations:

Number of plaques (pfu) in 15 µl of sample or control was calculated as follows:

$$\text{Pfu} = \frac{N_1 + N_2 + N_3}{3} \times \frac{150 \mu\text{l}}{100 \mu\text{l}} \times \text{dilution factor}$$

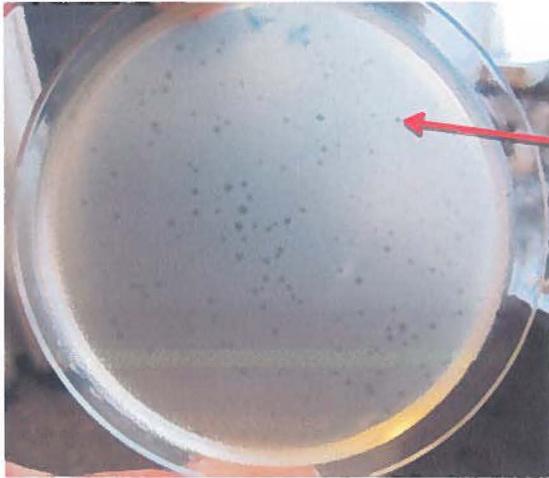
Here N₁, N₂, N₃ : number of plaques (pfu) from three separate tests.

150 µl : volume of diluted viruses (15 µl viruses + 135 µl H₂O).

100 µl : volume taken out to do plaque assays.

To evaluate the efficiency of the laser irradiation in inactivating viruses, the virus infectivity that remains after laser treatment was calculated as follows:

$$\text{Remaining virus infectivity (\%)} = \frac{\text{pfu/ml of treated sample}}{\text{pfu/ml of untreated control}} \times 100$$



A transparent plaque on the bacterial lawn

Figure 35. Example of a titration plate. The plaques are transparent circles on the white bacterial lawn.

RESULTS

1. Inactivation of M13 and MS2 with the 1 kHz laser and without focus

There is no significant change in the temperature of the samples (maximum temperature is 28°C) following inactivation. The samples and controls were diluted 10-fold to the dilution where the titer was 10^2 pfu. The number of plaques for M13 without laser irradiation was 107 ± 6 while plaque number after irradiation was 108 ± 7 . For bacteriophage MS2, plaque count for control is 111 ± 14 whereas that for laser-treated samples is 112 ± 14 . As a result, there was no inactivation of M13 or MS2 in this condition.

2. Inactivation of M13 with the 10 Hz laser and focus

2.1. Impact of time on virus inactivation

Bacteriophages M13 were irradiated by femtosecond laser with energy of 20 mJ, pulse width of 40 fs and at a wavelength of either 800 nm or 400 nm for 12 minutes, 24 minutes or 60 minutes. The results indicate a dependence of virus inactivation on the laser irradiation time at both wavelengths. As shown in Figure 36, virus infectivity was reduced with an increase in irradiation time. Although almost all viruses were inactivated for both lasers after 60 minutes, the 400 nm laser is more efficient than the 800 nm laser. For example, after 12 minutes of laser irradiation, only 0.11 % of viruses are still infectious after being irradiated with the laser at 400 nm, while up to 78% of the viruses retain their infectivity following treatment with 800 nm laser.

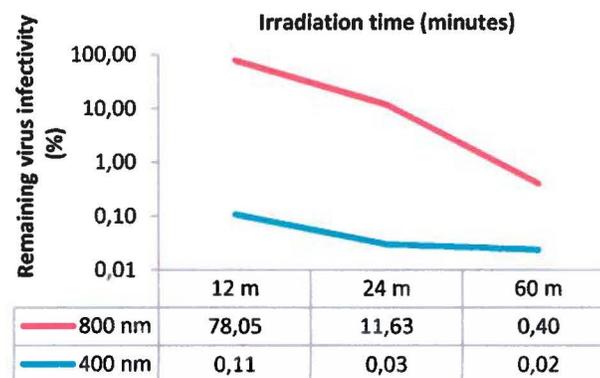


Figure 36. Impact of time on virus inactivation. The lasers are 20 mJ, 40 fs and at either 800 nm (red) or 400 nm (blue). The x-axis is irradiation time in minutes. The y-axis is remaining virus infectivity (%) of which the values are listed in the table below the graph. The standard deviations of these experiments are very small, ranging from 5×10^{-5} to 4×10^{-4} .

2.2. Impact of energy on virus inactivation

40 fs lasers at either 800 nm or 400 nm wavelengths were used to inactivate M13 during 24 minutes or 60 minutes, while the energy was changed to be 20 mJ, 15 mJ, 10 mJ or 5 mJ per pulse.

For the case of the 800 nm laser, we found that the lower the energy was, the fewer the viruses were inactivated. As shown in Figure 37, remaining virus infectivity increased from 11.63 % to 92.15 % when the laser energy decreased from 20 mJ to 5 mJ after 24 minutes of inactivation. The same trend can be seen for the viruses inactivated for 60 minutes, even though the virus infectivity was much smaller, for instance, only 39.69 % with 5 mJ laser.

On the other hand, the results of virus inactivation using the 400 nm laser did not show such a clear tendency (Figure 37). In fact, when the energy was reduced from 15 mJ to 10 mJ, the virus infectivity was the same after 24 minutes (0.01%) or even decreased after 60 minutes (0.04% for 15 mJ and 0.02% for 10 mJ). This is possibly an error and hence, repeat of this experiment needs to be done. Besides, the fact that the virus infectivity of samples treated with 5 mJ laser was higher than those with 20 mJ laser both after 24 minutes and 60 minutes suggests that virus irradiation also reduced following decrease in energy of the laser.

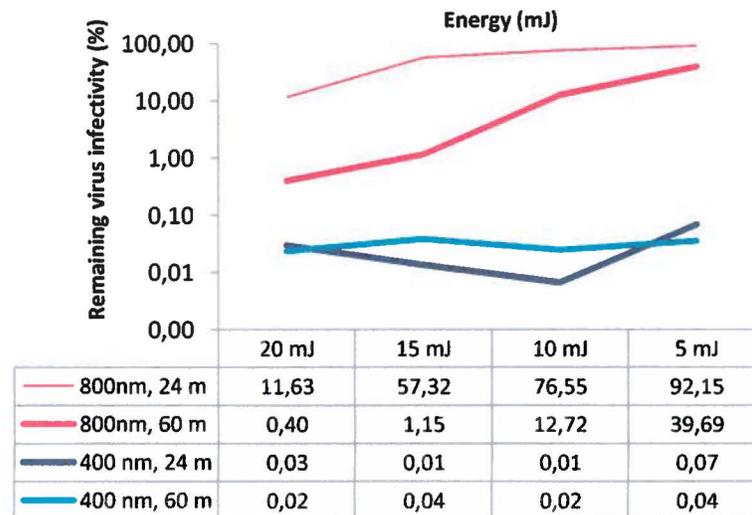


Figure 37. Impact of energy on virus inactivation. With the pulse width of 40 fs and at either 400 nm or 800 nm, M13 were inactivated with one of these energies: 20 mJ, 15 mJ, 10 mJ and 5 mJ. In the graph, the x-axis is the energy (mJ) while the y-axis is remaining virus infectivity (%) of which the values are listed in the table at the bottom. The standard deviations of these results are very small, from 2×10^{-6} to 4×10^{-5} .

2.3. Impact of pulse width on virus inactivation

To see whether changing the pulse width of the laser could affect virus inactivation, M13 bacteriophages were treated with lasers (800 nm or 400 nm) at different pulse width, ranging from 40 fs to 1500 fs. The energy for the experiments at 800 nm was always 20 mJ, whereas those at 400 nm decreased with an increase in the pulse width (Table 6).

For the 800 nm laser, virus infectivity was reduced when the pulse width was increased from 40 fs to 100 fs. In general, when the pulse width of the laser was increased, fewer viruses were inactivated (Figure 38). For example, after 60 minutes of irradiation at 800 nm, the 40 fs laser inactivated almost all M13, while treatment with the laser at 1250 fs led to no virus inactivation. However, there were exceptions. For instance, the 800 nm laser with 1500 fs pulse inactivated more viruses than the 1250 fs one. The same can be seen for the 400 nm laser. Therefore, this trend seems to be uncertain. Nevertheless, in the 400 nm as well as the 800 nm laser, the 40 fs laser pulse is the most efficient for virus inactivation. Even though increase in laser pulse duration rendered the inactivation less efficient, the virus inactivation by the 400 nm laser was still significant (remaining virus infectivity of only 2.3%) for the pulse width of 1500 fs.

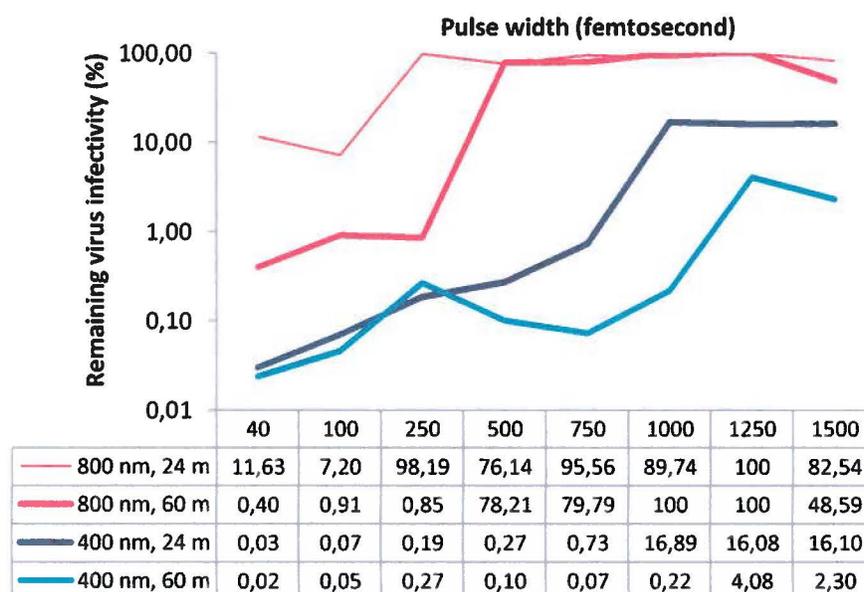


Figure 38. Impact of laser pulse width on virus inactivation. The energy of the laser was kept at 20 mJ, the wavelength is either 800 nm or 400 nm and the time is 24 minutes or 60 minutes while the laser pulse width (the x-axis) was changed. The y-axis is remaining virus infectivity (%). The standard deviations of these experiments are very small ranging from 8×10^{-5} to 3×10^{-3} .

2.4. Virus inactivation with both 400 nm and 800 nm laser

Lasers at both 800 nm and 400 nm wavelengths were used simultaneously to test whether there is any combined effect on virus inactivation. Measuring the energy gave an equal value of 10 mJ for each wavelength. We compared the remaining virus infectivity of three experiments, with the same energy (10 mJ), the same pulse width (40 fs) and the same irradiation time (24 minutes or 60 minutes). As shown in Figure 39, although two-colour irradiation (400 nm & 800 nm) inactivates more viral particles than 800 nm irradiation, it is still less effective than 400 nm irradiation. Take, for example, the 60 minute inactivation: 7.9 % of M13 treated with 400 nm & 800 nm laser retained their infectivity, while most viruses were inactivated with the 400 nm laser, and up to 12.72 % of the viruses treated with 800 nm laser are still infectious.

In addition, lasers with different pulse durations were tested with the two-colour irradiation. It is interesting that unlike one-colour irradiation (Figure 40), more viruses were inactivated at 100 fs and 250 fs, with remaining virus infectivity of 0.03 % and 0.21 %, respectively. However, much fewer viruses were inactivated with larger pulse width.

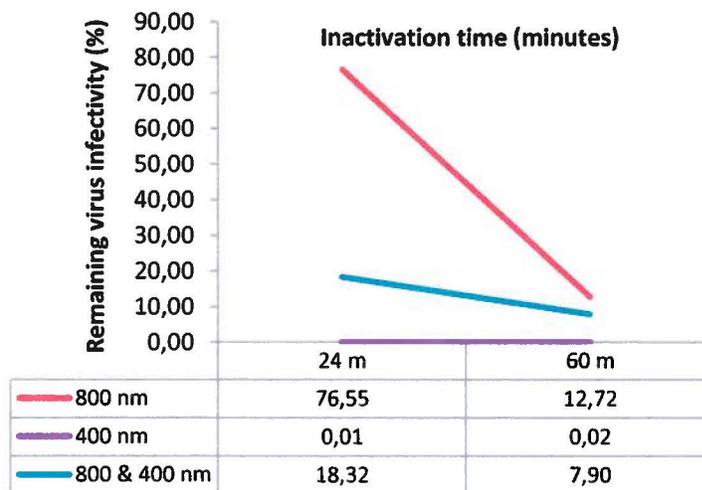


Figure 39. Comparison of remaining virus infectivity between one- and two-colour irradiation. For all the lasers, the energy is 10 mJ and the pulse width is 40 fs. Two durations were included which are 24 minutes and 60 minutes (x-axis). The y-axis is remaining virus infectivity in percentage. The standard deviations of these experiments are very small, ranging from 2×10^{-6} to 2×10^{-3} .

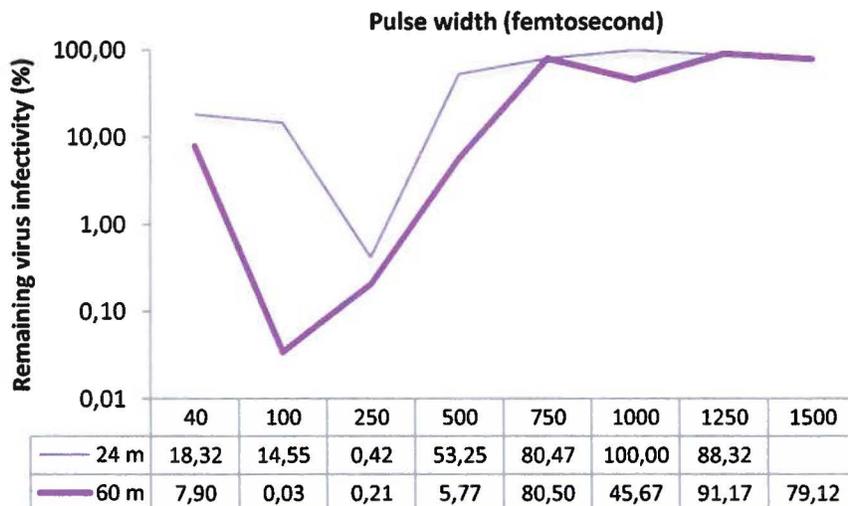


Figure 40. Virus inactivation with both 800 nm and 400 nm laser. Different pulse durations were tested for irradiation time of 24 minutes or 60 minutes. In the graph, the x-axis is the pulse duration (femtosecond) and the y-axis is the remaining virus infectivity (%). The standard deviations of these results are very small, from 3×10^{-4} to 2×10^{-3} .

DISCUSSION

We have seen that using the 800 nm laser with focus inactivated viruses after 1 hour, whereas the 800 nm laser without focus did not, which suggests the importance of focusing. Although the latter laser (57 GW) is less powerful than the other (0.5 TW), it is still much more powerful than the laser used by Tsen *et al.* (71) where virus inactivation occurred. However, the experiments of that group also required focus. Therefore, it is likely to be essential to focus the laser beam to the sample area to acquire significant virus inactivation.

For single colour irradiation, it is clear that the duration of laser irradiation determines its efficiency. Although not all viruses were inactivated after one hour, only very small amount of M13 particles are still infectious. As a result, maybe just a small increase in the time of inactivation (maximum half an hour) can be sufficient to achieve 100 % inactivation for both wavelengths. In the tests for the role of energy, a clear trend can be observed for the 800 nm laser, of which the outcome of the inactivation reduced with the decrease of laser energy. Yet there were some discrepancies in those for the 400 nm laser. This can be the consequence of errors in pipetting. Nevertheless, the differences in the inactivation for the tested maximum energy (20 mJ) and for the minimum energy (4 mJ), in addition to the clear trend of the 800 nm laser can be enough to say that the higher the energy of the laser, the more efficient the virus inactivation. Similarly, experiments to test a variety of pulse duration, also gave varying results. Nonetheless, it could possibly be due to pipetting errors. The majority of the results suggested a conclusion that the shorter the pulse duration, the more efficient the virus inactivation. This corresponds well with the theory that the intensity of the laser increases when its pulse duration becomes shorter.

For two-colour irradiation, a reduction in the virus inactivation capacity may be explained by the interference of the 800 nm laser (lower activity) with the 400 nm laser (higher activity). This might also be the reason why virus inactivation seems to be better with a longer pulse duration of somewhere between 100 fs and 250 fs.

For the moment, we do not have enough information from the experiments to pin down the mechanism for virus inactivation. However, there was only a minimal increase in temperature of the samples after inactivation, with a maximum temperature of 38°C being observed. Therefore the viruses were inactivated not because of the heat since normally these viruses live in the gut flora at 37°C. On the other hand, impulsive stimulated Raman scattering

would be another potential mechanism. The femtosecond laser creates a large spectrum of vibrations and one of which can associate with the Raman vibration of M13 particles, resulting in excitation of these viruses. Virus excitations were amplified up to a high level, where the weak links on viral capsid were broken, leading to damage of the capsid and inactivation of viral particles.

CONCLUSION

We used two femtosecond lasers, one with a repetition rate of 1 kHz and without focusing, and the other 10 Hz and with focusing. The lack of virus inactivation with the first laser suggests the important role of focusing the laser beam to the sample area.

For the 10 Hz laser, different experiments were performed to inactivate bacteriophage M13 in conditions varying in wavelength, time and pulse width. We found that the laser at 400 nm gave the best inactivation (almost 100%) after only 12 minutes of irradiation with the energy of 20 mJ and the pulse width of 40 fs. The laser at 800 nm is less intense, but after 60 minutes, almost all viruses were inactivated under the same conditions of energy and pulse width of either 40 fs or 100 fs. The decrease in energy as well as the increase in pulse width of the laser appears to reduce the efficiency of the virus inactivation. The combination of the two wavelengths could not improve inactivation with 40 fs pulse width laser although it is better than the 800 nm inactivation. However, good inactivation was observed when using the pulse duration of 100 fs and 250 fs, suggesting maybe the pulse width somewhere in this range can facilitate virus inactivation for two-colour irradiations.

These results are only preliminary and thus need to be confirmed with the same viruses as well as others, especially pathogenic viruses and even other pathogens. Experiments with beneficial components like human cells and proteins are also necessary. According to the group of Tsen, the power density of $\leq 10 \text{ GW cm}^{-2}$ can be used to selectively inactivate pathogens while more than that would injure cells like red blood cells. However the power density in our experiments is much higher than those of this group (0.6 TW/cm^2). Whether this high intensity can cause adverse effects remains to be studied.

Because of its high efficiency, safety and cost-effectiveness, once thoroughly studied, this is a very promising method to selectively eliminate pathogens in biological products such as human whole blood or blood components.

REFERENCES

1. **Abd-Alla, A., F. X. Jousset, Y. Li, G. Fediere, F. Cousserans, and M. Bergoin.** 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**:790-7.
2. **Alter, H. J.** 2008. Pathogen reduction: a precautionary principle paradigm. *Transfus Med Rev* **22**:97-102.
3. **Altschul, S. F., T. L. Madden, A. A. Schaffer, J. Zhang, Z. Zhang, W. Miller, and D. J. Lipman.** 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25**:3389-402.
4. **Altschul, S. F., J. C. Wootton, E. M. Gertz, R. Agarwala, A. Morgulis, A. A. Schaffer, and Y. K. Yu.** 2005. Protein database searches using compositionally adjusted substitution matrices. *Febs J* **272**:5101-9.
5. **Anderson, L. J., and N. S. Young.** 1997. Human Parvovirus B19, p. 105-119. *In* L. J. A. a. N. Young (ed.), *Human parvovirus B19: pathogenesis of disease*. Karger, Basel.
6. **Benjamin, R. J., J. McCullough, P. D. Mintz, E. Snyder, W. D. Spotnitz, R. J. Rizzo, D. Wages, J. S. Lin, L. Wood, and L. Corash.** 2005. Therapeutic efficacy and safety of red blood cells treated with a chemical process (S-303) for pathogen inactivation: a Phase III clinical trial in cardiac surgery patients. *Transfusion* **45**:1739-49.
7. **Benkert, P., M. Biasini, and T. Schwede.** 2011. Toward the estimation of the absolute quality of individual protein structure models. *Bioinformatics* **27**:343-50.
8. **Brown, K. E., S. M. Anderson, and N. S. Young.** 1993. Erythrocyte P antigen: cellular receptor for B19 parvovirus. *Science* **262**:114-7.
9. **Bryant, B. J., and H. G. Klein.** 2007. Pathogen inactivation: the definitive safeguard for the blood supply. *Arch Pathol Lab Med* **131**:719-33.
10. **Burnouf, T., and M. Radosevich.** 2003. Nanofiltration of plasma-derived biopharmaceutical products. *Haemophilia* **9**:24-37.
11. **Burnouf, T., M. Radosevich, H. Goubran, and H. Willkommen.** 2005. Place of nanofiltration for assuring viral safety of biologicals. *Current Nanoscience* **1**:189-201.
12. **Capinera, J. L.** 1999. Cabbage Looper, *Trichoplusia ni* (Hübner) (Insecta: Lepidoptera: Noctuidae). Gainesville: University of Florida Institute of Food and Agricultural Sciences.
13. **Chao, Y.-C., S. Y. Young Iii, K. S. Kim, and H. A. Scott.** 1985. A newly isolated denonucleosis virus from *Pseudoplusia includens* (Lepidoptera: Noctuidae). *Journal of Invertebrate Pathology* **46**:70-82.
14. **Cohn, E., L. E. Strong, W. L. Hughes, D. Mulford, J. Ashworth, M. Melin, and H. Taylor.** 1946. Preparation and Properties of Serum and Plasma Proteins. IV. A System for the Separation into Fractions of the Protein and Lipoprotein Components of Biological Tissues and Fluids1a, b, c, d. *Journal of the American Chemical Society* **68**:459-75.

15. **Corash, L.** 2001. Helinx technology for inactivation of infectious pathogens and leukocytes in labile blood components: from theory to clinical application. *Transfusion and Apheresis Science* **25**:179-81.
16. **Corash, L.** 2001. Inactivation of infectious pathogens in labile blood components: meeting the challenge. *Transfusion clinique et biologique* **8**:138-45.
17. **Cotmore, S. F., and P. Tattersall.** 1987. The autonomously replicating parvoviruses of vertebrates. *Adv Virus Res* **33**:91-174.
18. **Cotmore, S. F., Tattersall, P.** 2006. Structure and organization of the viral genome, p. 73-94. *In* S. F. C. J.R. Kerr, M.E. Bloom, R.M. Linden and C.R. Parrish (ed.), *Parvoviruses*. Hodder Arnold, London.
19. **Cotmore, S. F. a. T., P.** 2006. A rolling-hairpin strategy: basic mechanisms of DNA replication in the parvoviruses, p. 171. *In* S. F. C. J.R. Kerr, M.E. Bloom, R.M. Linden and C.R. Parrish (ed.), *Parvoviruses*. Hodder Arnold, London.
20. **Dardare, N., and M. S. Platz.** 2002. Binding Affinities of Commonly Employed Sensitizers of Viral Inactivation. *Photochemistry and photobiology* **75**:561-4.
21. **de Castro, E., C. J. Sigrist, A. Gattiker, V. Bulliard, P. S. Langendijk-Genevaux, E. Gasteiger, A. Bairoch, and N. Hulo.** 2006. ScanProsite: detection of PROSITE signature matches and ProRule-associated functional and structural residues in proteins. *Nucleic Acids Res* **34**:W362-5.
22. **Ding, C., M. Urabe, M. Bergoin, and R. M. Kotin.** 2002. Biochemical characterization of Junonia coenia densovirus nonstructural protein NS-1. *J Virol* **76**:338-45.
23. **Dodd, R. Y., E. P. t. Notari, and S. L. Stramer.** 2002. Current prevalence and incidence of infectious disease markers and estimated window-period risk in the American Red Cross blood donor population. *Transfusion* **42**:975-9.
24. **Dumas, B., M. Jourdan, A. M. Pascaud, and M. Bergoin.** 1992. Complete nucleotide sequence of the cloned infectious genome of Junonia coenia densovirus reveals an organization unique among parvoviruses. *Virology* **191**:202-22.
25. **Fediere, G., M. El-Far, Y. Li, M. Bergoin, and P. Tijssen.** 2004. Expression strategy of denonucleosis virus from Mythimna loreyi. *Virology* **320**:181-9.
26. **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.
27. **Ford, L.** 2003. Estimate of the vibrational frequencies of spherical virus particles. *Physical Review E* **67**:051924.
28. **Fryer, J. F., E. Delwart, F. Bernardin, P. W. Tuke, V. V. Lukashov, and S. A. Baylis.** 2007. Analysis of two human parvovirus PARV4 genotypes identified in human plasma for fractionation. *J Gen Virol* **88**:2162-7.
29. **Ghosh, D.** 1993. Status of the transcription factors database (TFD). *Nucleic Acids Res* **21**:3117-8.
30. **Girod, A., C. E. Wobus, Z. Zadori, M. Ried, K. Leike, P. Tijssen, J. A. Kleinschmidt, and M. Hallek.** 2002. The VP1 capsid protein of adeno-associated virus

- type 2 is carrying a phospholipase A2 domain required for virus infectivity. *J Gen Virol* **83**:973-8.
31. **Godiska, R., D. Mead, V. Dhodda, C. Wu, R. Hochstein, A. Karsi, K. Usdin, A. Entezam, and N. Ravin.** 2010. Linear plasmid vector for cloning of repetitive or unstable sequences in *Escherichia coli*. *Nucleic Acids Res* **38**:29.
 32. **Hebert, B., J. Bergeron, E. F. Potworowski, and P. Tijssen.** 1993. Increased PCR sensitivity by using paraffin wax as a reaction mix overlay. *Mol Cell Probes* **7**:249-52.
 33. **Hengen, P. N.** 1997. Optimizing multiplex and LA-PCR with betaine. *Trends Biochem Sci* **22**:225-6.
 34. **Huang, X., and A. Madan.** 1999. CAP3: A DNA sequence assembly program. *Genome Res* **9**:868-77.
 35. **Huynh, O. T. H., H. T. Pham, Q. Yu, and P. Tijssen.** 2012. Pseudoplusia includens Densovirus Genome Organization and Expression Strategy. *J Virol* **86**:13127-8.
 36. **Jung, A., S. Ruckert, P. Frank, T. Brabletz, and T. Kirchner.** 2002. 7-Deaza-2'-deoxyguanosine allows PCR and sequencing reactions from CpG islands. *Mol Pathol* **55**:55-7.
 37. **Kang, J., M. S. Lee, and D. G. Gorenstein.** 2005. The enhancement of PCR amplification of a random sequence DNA library by DMSO and betaine: application to in vitro combinatorial selection of aptamers. *J Biochem Biophys Methods* **64**:147-51.
 38. **Kiefer, F., K. Arnold, M. Kunzli, L. Bordoli, and T. Schwede.** 2009. The SWISS-MODEL Repository and associated resources. *Nucleic Acids Res* **37**:18.
 39. **Kilham, L., and L. J. Olivier.** 1959. A latent virus of rats isolated in tissue culture. *Virology* **7**:428-37.
 40. **Klein, H., R. Dodd, W. Dzik, N. Luban, P. Ness, P. Pisciotto, P. Schiff, and E. Snyder.** 2002. Current status of solvent/detergent-treated frozen plasma. *Transfusion* **38**:102-7.
 41. **Lezin, G., Y. Kosaka, H. J. Yost, M. R. Kuehn, and L. Brunelli.** 2011. A one-step miniprep for the isolation of plasmid DNA and lambda phage particles. *PLoS One* **6**:15.
 42. **Li, Y., Z. Zadori, H. Bando, R. Dubuc, G. Fediere, J. Szelei, and P. Tijssen.** 2001. Genome organization of the densovirus from *Bombyx mori* (BmDENV-1) and enzyme activity of its capsid. *J Gen Virol* **82**:2821-5.
 43. **Lin, L., D. Cook, G. Wieseahn, R. Alfonso, B. Behrman, G. Cimino, L. Corten, P. Damonte, R. Dikeman, and K. Dupuis.** 2003. Photochemical inactivation of viruses and bacteria in platelet concentrates by use of a novel psoralen and long-wavelength ultraviolet light. *Transfusion* **37**:423-35.
 44. **Marchler-Bauer, A., S. Lu, J. B. Anderson, F. Chitsaz, M. K. Derbyshire, C. DeWeese-Scott, J. H. Fong, L. Y. Geer, R. C. Geer, N. R. Gonzales, M. Gwadz, D. I. Hurwitz, J. D. Jackson, Z. Ke, C. J. Lanczycki, F. Lu, G. H. Marchler, M. Mullokandov, M. V. Omelchenko, C. L. Robertson, J. S. Song, N. Thanki, R. A. Yamashita, D. Zhang, N. Zhang, C. Zheng, and S. H. Bryant.** 2011. CDD: a

- Conserved Domain Database for the functional annotation of proteins. *Nucleic Acids Res* **39**:24.
45. **Mintz, P. D., A. Neff, M. MacKenzie, L. T. Goodnough, C. Hillyer, C. Kessler, K. McCrae, J. E. Menitove, B. S. Skikne, and L. Damon.** 2006. A randomized, controlled Phase III trial of therapeutic plasma exchange with fresh-frozen plasma (FFP) prepared with amotosalen and ultraviolet A light compared to untreated FFP in thrombotic thrombocytopenic purpura. *Transfusion* **46**:1693-704.
 46. **Murphy, S., E. Snyder, R. Cable, S. J. Slichter, R. G. Strauss, J. McCullough, J. S. Lin, L. Corash, and M. G. Conlan.** 2005. Platelet dose consistency and its effect on the number of platelet transfusions for support of thrombocytopenia: an analysis of the SPRINT trial of platelets photochemically treated with amotosalen HCl and ultraviolet A light. *Transfusion* **46**:24-33.
 47. **Musso, M., R. Bocciardi, S. Parodi, R. Ravazzolo, and I. Ceccherini.** 2006. Betaine, dimethyl sulfoxide, and 7-deaza-dGTP, a powerful mixture for amplification of GC-rich DNA sequences. *J Mol Diagn* **8**:544-50.
 48. **Papadopoulos, J. S., and R. Agarwala.** 2007. COBALT: constraint-based alignment tool for multiple protein sequences. *Bioinformatics* **23**:1073-9.
 49. **Parker, J. S., W. J. Murphy, D. Wang, S. J. O'Brien, and C. R. Parrish.** 2001. Canine and feline parvoviruses can use human or feline transferrin receptors to bind, enter, and infect cells. *J Virol* **75**:3896-902.
 50. **Pettersen, E. F., T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng, and T. E. Ferrin.** 2004. UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem* **25**:1605-12.
 51. **Pomp, D., and J. F. Medrano.** 1991. Organic solvents as facilitators of polymerase chain reaction. *Biotechniques* **10**:58-9.
 52. **Purmal, A., C. R. Valeri, W. Dziki, L. Pivacek, G. Ragno, A. Lazo, and J. Chapman.** 2002. Process for the preparation of pathogen-inactivated RBC concentrates by using PEN110 chemistry: preclinical studies. *Transfusion* **42**:139-45.
 53. **Ranu, R. S.** 1994. Relief of DNA polymerase stop(s) due to severity of secondary structure of single-stranded DNA template during DNA sequencing. *Anal Biochem* **217**:158-61.
 54. **Rees, W. A., T. D. Yager, J. Korte, and P. H. von Hippel.** 1993. Betaine can eliminate the base pair composition dependence of DNA melting. *Biochemistry* **32**:137-44.
 55. **Reese, M. G.** 2001. Application of a time-delay neural network to promoter annotation in the *Drosophila melanogaster* genome. *Comput Chem* **26**:51-6.
 56. **Rubinstein, A. I., and D. B. Rubinstein.** Inability of solvent-detergent (S-D) treated factor VIII concentrate to inactivate parvoviruses and non-lipid enveloped non-A, non-B hepatitis virus in factor VIII concentrate: advantages to using sterilizing 100 degrees C dry heat treatment. *Am J Hematol.* 1990 Oct;**35**(2):142.
 57. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1992. DNA sequencing with chain-terminating inhibitors. 1977. *Biotechnology* **24**:104-8.

58. **Sigrist, C. J., L. Cerutti, E. de Castro, P. S. Langendijk-Genevaux, V. Bulliard, A. Bairoch, and N. Hulo.** 2010. PROSITE, a protein domain database for functional characterization and annotation. *Nucleic Acids Res* **38**:25.
59. **Simpson, A. A., P. R. Chipman, T. S. Baker, P. Tijssen, and M. G. Rossmann.** 1998. The structure of an insect parvovirus (*Galleria mellonella* densovirus) at 3.7 Å resolution. *Structure* **6**:1355-67.
60. **Solovyev, V. V., and I. A. Shakhmuradov.** 2003. PromH: Promoters identification using orthologous genomic sequences. *Nucleic Acids Res* **31**:3540-5.
61. **Sowemimo-Coker, S., R. Kasczak, A. Kim, F. Andrade, S. Pesci, C. Meeker, R. Carp, and P. Brown.** 2005. Removal of exogenous (spiked) and endogenous prion infectivity from red cells with a new prototype of leukoreduction filter. *Transfusion* **45**:1839-44.
62. **Strauss, J. H., Jr., and R. L. Sinsheimer.** 1963. Purification and properties of bacteriophage MS2 and of its ribonucleic acid. *J Mol Biol* **7**:43-54.
63. **Summerford, C., and R. J. Samulski.** 1998. Membrane-associated heparan sulfate proteoglycan is a receptor for adeno-associated virus type 2 virions. *J Virol* **72**:1438-45.
64. **Szelei, J., K. Liu, Y. Li, S. Fernandes, and P. Tijssen.** 2010. Parvovirus 4-like virus in blood products. *Emerg Infect Dis* **16**:561-4.
65. **Tattersall, P.** 2006. The evolution of parvovirus taxonomy, p. 5-14. *In* M. E. B. S. F. C. J.R. Kerr, R.M. Linden and C.R. Parrish (ed.), *Parvoviruses*. Hodder Arnold, London.
66. **Tijssen, P., Bando, H., Li, Y., Jousset, F.-X., Zadori, F.G., El-Far, M.Z., Szelei, J. and Bergoin, M.** 2006. Evolution of densovirus, p. 55–68. *In* S. F. C. J.R. Kerr, M.E. Bloom, R.M. Linden and C.R. Parrish (ed.), *Parvoviruses*. Hodder Arnold, London.
67. **Tijssen, P., Y. Li, M. El-Far, J. Szelei, M. Letarte, and Z. Zadori.** 2003. Organization and expression strategy of the ambisense genome of densovirus of *Galleria mellonella*. *J Virol* **77**:10357-65.
68. **Tijssen, P., M. Agbandje-McKenna, J. M. Almendral, M. Bergoin, T. W. Flegel, K. Hedman, J. A. Kleinschmidt, Y. Li, D. J. Pintel and P. Tattersall.** 2011. Parvoviridae, p. 375-395. *In* M. J. A. A. M. Q. King, E. Carstens, and E. J. Lefkowitz (ed.), *Virus Taxonomy: classification and nomenclature of viruses: Ninth Report of the International Committee on Taxonomy of Viruses*. Elsevier Science, San Diego.
69. **Tsen, K., E. C. Dykeman, O. F. Sankey, S. W. D. Tsen, N. T. Lin, and J. G. Kiang.** 2006. Raman scattering studies of the low-frequency vibrational modes of bacteriophage M13 in water—observation of an axial torsion mode. *Nanotechnology* **17**:5474.
70. **Tsen, K., S. W. D. Tsen, C. L. Chang, C. F. Hung, T. Wu, and J. G. Kiang.** 2007. Inactivation of viruses with a very low power visible femtosecond laser. *Journal of Physics: Condensed Matter* **19**:322102.
71. **Tsen, K. T., S. W. D. Tsen, Q. Fu, S. M. Lindsay, K. Kibler, B. Jacobs, T. C. Wu, B. Karanam, S. Jagu, and R. B. S. Roden.** 2009. Photonic approach to the selective inactivation of viruses with a near-infrared subpicosecond fiber laser. *Journal of Biomedical Optics* **14**:064042-7.

72. **Viswanathan, V. K., K. Krcmarik, and N. P. Cianciotto.** Template secondary structure promotes polymerase jumping during PCR amplification. *Biotechniques*. 1999 Sep;27(3):508-11.
73. **Wagner, S. J.** 2002. Virus inactivation in blood components by photoactive phenothiazine dyes. *Transfusion medicine reviews* 16:61.
74. **Wainwright, M., H. Mohr, and W. H. Walker.** 2007. Phenothiazinium derivatives for pathogen inactivation in blood products. *Journal of Photochemistry and Photobiology B: Biology* 86:45-58.
75. **Walters, R. W., S. M. Yi, S. Keshavjee, K. E. Brown, M. J. Welsh, J. A. Chiorini, and J. Zabner.** 2001. Binding of adeno-associated virus type 5 to 2,3-linked sialic acid is required for gene transfer. *J Biol Chem* 276:20610-6.
76. **Wang, X. S., K. Qing, S. Ponnazhagan, and A. Srivastava.** 1997. Adeno-associated virus type 2 DNA replication in vivo: mutation analyses of the D sequence in viral inverted terminal repeats. *J Virol* 71:3077-82.
77. **Weitzman, M. D.** 2006. The parvovirus life cycle: an introduction to molecular interactions important for infection, p. 143-156. *In* S. F. C. J.R. Kerr, M.E. Bloom, R.M. Linden and C.R. Parrish (ed.), *Parvoviruses*. Hodder Arnold, London.
78. **Williamson, L. M., R. Cardigan, and C. V. Prowse.** 2003. Methylene blue-treated fresh-frozen plasma: what is its contribution to blood safety? *Transfusion* 43:1322-9.
79. **Willwand, K., and B. Hirt.** 1991. The minute virus of mice capsid specifically recognizes the 3' hairpin structure of the viral replicative-form DNA: mapping of the binding site by hydroxyl radical footprinting. *J Virol* 65:4629-35.
80. **Yan, Y. X., E. B. Gamble, and K. A. Nelson.** 1985. Impulsive stimulated scattering: General importance in femtosecond laser pulse interactions with matter, and spectroscopic applications. *The Journal of chemical physics* 83:5391-9.
81. **Zadori, Z., J. Szelei, M. C. Lacoste, Y. Li, S. Gariepy, P. Raymond, M. Allaire, I. R. Nabi, and P. Tijssen.** 2001. A viral phospholipase A2 is required for parvovirus infectivity. *Dev Cell* 1:291-302.
82. **Zhang, Z., S. Schwartz, L. Wagner, and W. Miller.** 2000. A greedy algorithm for aligning DNA sequences. *J Comput Biol* 7:203-14.
83. **Zuker, M.** 2003. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res* 31:3406-15.

