UNIVERSIDADE DE BRASÍLIA
INSTITUTO DE CIÊNCIAS BIOLÓGICAS
DEPARTAMENTO DE BIOLOGIA CELULAR
PÓS-GRADUAÇÃO EM BIOLOGIA MOLECULAR

Genômica, evolução e caracterização funcional de genes de baculovírus

DANIEL M. P. ARDISSON-ARAÚJO

Orientador: Dr. Bergmann Morais Ribeiro
Co-orientador: Dr. Fernando Lucas Melo
Orientador estrangeiro: Dr. Rollie J. Clem

Brasília, 2015.

UNIVERSIDADE DE BRASÍLIA
INSTITUTO DE CIÊNCIAS BIOLÓGICAS
DEPARTAMENTO DE BIOLOGIA CELULAR
PÓS-GRADUAÇÃO EM BIOLOGIA MOLECULAR

Genômica, evolução e caracterização funcional de genes de baculovírus

DANIEL M. P. ARDISSON-ARAÚJO

Orientador: Dr. Bergmann Morais Ribeiro
Co-orientador: Dr. Fernando Lucas Melo
Orientador estrangeiro: Dr. Rollie J. Clem

Tese apresentada ao Programa de Pós-Graduação em Ciências Biológicas – Biologia Molecular, do Departamento de Biologia Celular, do Instituto de Ciências Biológicas da Universidade de Brasília como parte dos requisitos para obtenção do título de Doutor em Biologia Molecular.

Brasília, 2015.

DANIEL M. P. ARDISSON-ARAÚJO

Genômica, evolução e caracterização funcional de genes de baculovírus

Tese apresentada ao Programa de Pós-Graduação em Ciências Biológicas – Biologia Molecular, do Departamento de Biologia Celular, do Instituto de Ciências Biológicas da Universidade de Brasília como parte dos requisitos para obtenção do título de Doutor em Biologia Molecular.

Banca Examinadora:

_____

Prof. Dr. Bergmann Morais Ribeiro (Orientador) (CEL – UnB)

_____

Profa. Dra. Ildinete Silva-Pereira (CEL – UnB)

_____

Profa. Erna Geessien Kroon (ICB/UFMG)

_____

Prof. Dr. Jônatas Santos Abrahão (ICB/UFMG)

_____

Prof. Dr. Ricardo Henrique Kruger (CEL/UnB)

“Até onde posso, vou deixando o melhor de mim...
Se alguém não viu, foi porque não me sentiu com o coração.”
Clarice Lispector

The greatest enemy of knowledge is not ignorance,
it is the illusion of knowledge”.
Stephen Hawking

“Nothing in biology makes sense except in the light of evolution”.
‘Nothing in life makes sense except in the light of changing’ (paráfrase).
Theodosius Dobzhansky

“E conhecereis a verdade e a verdade vos libertará”.
João 8:32

"Y las verdades se suceden en distintas épocas.
No existe solamente una verdad."
Mercedes Sosa

A minha família

Ao prof. Bergmann M. Ribeiro

Agradecimento

Agradeço a todos que de alguma forma colaboraram para a elaboração desta tese, direta ou indiretamente. Agradeço à minha família: Moza, mãe e mãezona. Agradeço aos meus amigos de perto e de longe, cujo tropismo no coração é bem desenhado, alguns superficiais e agudos, outros profundos e crônicos. Agradeço aos professores que de alguma forma abriram caminho para esta história de amor profundo pelo conhecimento e pela busca de uma verdade, ainda que transiente. Agradeço em especial ao meu orientador de tantos anos, prof. Bergmann Morais Ribeiro. Agradeço ao meu co-orientador, Fernando Lucas de Melo. Agradeço ao meu orientador estrangeiro, prof. Rollie J. Clem por me receber durante meu período sanduíche. Agradeço aos componentes da banca de ambos qualificação e de defesa de tese. Agradeço ao Brasil, ao CNPq, à CAPES, à FAP/DF, à Universidade de Brasília e a Kansas State University.

Índice

Resumo

Baculovirus são vírus de DNA dupla-fita circular capazes de infectar oralmente o estágio larval de insetos. Atualmente, são usados para o controle biológico de insetos praga e como vetores de expressão de proteínas heterólogas. Pouco é sabido das bases moleculares da interação do vírus com o hospedeiro e de sua evolução. Os fatores limitantes estão associados ao número de genomas sequenciados bem como a restrição do cultivo *in vitro* de várias espécies virais. De fato, a base para o início de quaisquer estudos moleculares mais detalhados de novas espécies de baculovírus ou de isolados certamente se inicia com o sequenciamento do genoma completo e com o estudo de genes encontrados. Dessa forma, neste trabalho, vários genomas de baculovírus isolados no Brasil foram sequenciados e descritos. Sequenciamos e descrevemos baculovírus isolados do mandarová-da-mandicoca, da broca da cana-de-açúcar, do bicho da seda, da lagarta polífaga *Helicoverpa armigera*, do mandarová-do-mate entre outros. Concomitante à descrição do genoma, caracterizamos estruturalmente algumas espécies, avaliamos a taxa de mortalidade em situações controladas de infecção, bem como caracterizamos alguns genes que permitiram um entendimento evolutivo mais amplo das espécies descritas e de sua interação com o hospedeiro. Descrevemos o primeiro inibidor de serino protease de baculovírus capaz de bloquear a imunidade inata do inseto hospedeiro e causar proteção ao patógeno. Encontramos o primeiro betabaculovírus com uma proteína de fusão de envelope de alphabaculovírus, a gp64 e caracterizamos sua funcionalidade. Além disso, mostramos pela primeira vez o papel de genes envolvidos no metabolismo de nucleotídeo e sua capacidade de alterar o desempenho viral. Em conclusão, baculovírus apresentam plasticidade genômica com aquisições proeminentes de genes de vários organismos como outros vírus de insetos, bactérias e plantas. Além disso, perdas de genes ancestrais e duplicação são eventos recorrentes. Tanto a genômica quanto o estudo molecular básico de baculovírus tem contribuído para a compreensão de doenças associadas a humanos como câncer e doenças virais cujo agente etiológico apresenta genoma com DNA dupla-fita ou que infectam primariamente o intestino médio de insetos, como herpesvírus e arboviroses, respectivamente.

Palavras-chave: baculovírus, betabaculovírus, alphabaculovírus, genômica, evolução, transferência horizontal de genes.

Abstract

Baculoviruses are circular double-stranded DNA viruses that are orally infectious to larval stages of insects. Nowadays, they are used as biological control agents of agricultural and forest pests and as vector for heterologous protein expression. The understanding of both the molecular basis and the evolution of the virus/host interaction is scarce due to the few numbers of sequenced genomes and the restriction in cultivating several virus species *in vitro*. In fact, the beginning of any molecular study of new baculovirus species or isolates certainly pervades the whole genome sequencing. Therefore, in this work, several genomes of baculoviruses isolated in Brazil were sequenced and described. We sequenced and described baculoviruses isolated from subject cadavers of the cassava hornworm (*Erinnyis ello*), the sugar cane borer (*Diatraea saccharalis*), the silkworm (*Bombyx mori*), the bollworm (*Helicoverpa armigera*), and the mate hornworm (*Perigonia lusca*). Together with the genome description, we characterized structurally some species, evaluated the mortality in controlled infections, and characterized as well some genes to better understand the novel species and their interaction with the host. We described the first baculoviral serine protease inhibitor capable of blocking the insect immunity response and causing pathogen protection. We found the first betabaculovirus harboring an alphabaculovirus envelope fusion protein, a gp64 and we characterized its functionality. Furthermore, we have shown for the first time a role of genes related to nucleotide metabolism and it ability of altering the virus fitness. In conclusion, baculoviruses present genomic plasticity with great and recurrent acquisition of genes from several organisms including other insect viruses, bacteria, and plant. Moreover, ancestral gene losses and duplication are common events in baculovirus evolution. Both genomics and molecular biology of baculovirus have contributed to the comprehension of human-associated diseases such as cancer and viral whereas the etiologic agent presents dsDNA genome or infects primarily the insect midgut like herpexviruses and arboviruses, respectively.

Keywords: baculovirus, betabaculovirus, alphabaculovíius, genomics, evolution, horizontal gene transfer.

Capítulo 1. Introdução

1. Baculovírus

Baculovirus são vírus de DNA dupla-fita circular capazes de infectar oralmente o estágio larval de insetos das ordens Diptera (mosquitos da família Culicidae), Hymenoptera (larvas de vespa da família Diprionidae, que se comportam como lagartas) e Lepidoptera (mariposas e borboletas) (Rohrmann, 2013). No cenário mundial atual, baculovírus são poderosas ferramentas para o controle biológico de populações de insetos praga e vetores de expressão de proteínas heterólogas, além de apresentarem uso potencial como entregadores para terapia gênica (Summers, 2006; Ribeiro *et al.*, 2015).

O nome baculovírus, deriva do latim *baculo* que significa bastão, devido ao formato do nucleocapsideo viral (Rohrmann, 2013). Durante um ciclo infectivo completo, os vírus produzem dois fenótipos: (*i*) o vírion derivado de oclusão (ODV, do inglês 'occlusion-derived virion') que é responsável pela infecção oral e está ocluído num corpo cristalino proteico chamado de corpo de oclusão (OB, do inglês, 'occlusion body') e (*ii*) o vírion brotado (BV, do inglês, 'budded virion') responsável pelo espalhamento da infecção ao longo do corpo do inseto hospedeiro (Clem & Passarelli, 2013).

A rota de infecção do hospedeiro se inicia com a larva ingerindo alimentos (*e.g.* folha, ramos, frutos, caules, ou água no caso de larvas de mosquitos filtradores) contaminados por OBs. Os OBs atingem o intestino médio da larva e se dissolvem quando em contato com o pH alcalino do suco gástrico. Além dos vírions, a solubilização dos OBs libera enzimas que digerem a membrana peritrófica do lúmen intestinal, e permitem a

passagem das partículas infectivas em direção às células absortivas. ODVs infectam células colunares do intestino médio por fusão direta às microvilosidades e liberam nucleocapsídeo no citoplasma. O nucleocapsídeo é então direcionado por filamentos de actina para o núcleo, onde se desmonta e expõe o genoma viral para a maquinaria celular (Slack & Arif, 2007).

Inicialmente, durante a fase prococe da infecção, baculovírus manipulam a célula hospedeira causando o desligamento da expressão de proteínas da célula (Ooi & Miller, 1988). Toda a maquinaria celular fica a mercê do vírus, e trabalha a fim de produzir progênie viral durante a fase tardia da infecção. Depois de replicado, o genoma viral é montado em nucleocapsídeos e direcionado para a membrana da célula, de onde brotam como BVs. Os BVs espalham a infecção ao longo do corpo do inseto hospedeiro e estabelecem, dessa forma, a infecção secundária sistêmica. Depois da fase de produção de vírus brotados, a célula infectada ativa uma cascata de genes muito tardios virais responsáveis pela produção de ambos ODVs e OBs, encerrando assim o ciclo de infecção (Rohrmann, 2013).

A família *Baculoviridae* está agrupada em quatro gêneros, com base no alinhamento de 37 genes compartilhados (Jehle *et al.*, 2006). Este agrupamento converge com o espectro de hospedeiro e com características morfológicas dos OBs. Representantes do gênero *Alphabaculovirus* infectam insetos da ordem Lepidoptera e apresentam OBs poliédricos com tamanho de 800-2.000 nm. Estes podem ser agrupoados ainda em grupo I ou grupo II. A primeira sugestão de agrupamento ocorreu com base em análise filogenética da proteína formadora do corpo de oclusão, a poliedrina (Zonotto *et al.*, 1993). Posteriormente, foi observado que o tipo de proteína de fusão ao receptor celular

do fenótipo BV também era diferente de acordo com o grupo. Representantes do gênero *Betabaculovirus* infectam insetos também da ordem Lepidoptera, porém apresentam OBs com a forma de grânulos semelhantes a grãos de arroz com dimensões de 500 nm de altura e 200 nm de largura. Os gêneros *Gammabaculovirus* e *Deltabaculovirus* são infectivos a Hymenoptera e Diptera, respectivamente e ambos apresentam OBs poliédricos. Importante, baculovírus com OBs poliédricos são denominados de nucleopolyhedrovirus (NPVs) enquanto que aqueles com OBs granulares são chamados de granulovirus (GVs) e ambos os termos, antigamente reconhecidos como gêneros parafiléticos, são ainda usados na nomenclatura das espécies virais.

Quanto à anatomia dos vírions, ODVs e BVs apresentam nucleocapsídeos estruturalmente semelhantes entre si. Dessa forma, a principal diferença estrutural, composicional e funcional dos virions é gerada pelo envelope e por proteínas associadas (Braconi *et al.*, 2014). O envelope de BV apresenta uma região peplomérica responsável pela ligação ao receptor da célula hospedeira; cuja principal proteína de fusão de envelope (EFP, do inglês 'envelope fusion protein') é a proteína F em alphabaculovírus grupo II, betabaculovírus e deltabaculovírus ou sua análoga funcional adquirida posteriomente em alphabaculovirus grupo I, a proteína GP64 (Herniou & Jehle, 2007; Jehle *et al.*, 2006). As EFPs promovem endocitose adsortiva com receptores desconhecidos na superfície da célula hospedeira e, conforme maturação acídica do endossomo, sofrem modificação estrutural que permite fusão do envelope com a membrana do endossomo e liberação do nucleocapsídeo no citoplasma da célula (Wang *et al.*, 2014). Por outro lado, gammabaculovirus (baculovírus infectivos para hymenopteros) não codificam proteínas de envelope análogas à proteína F ou à GP64 em seu genoma, e dessa forma parecem não formar BVs durante o ciclo infectivo

completo (Rohrmann, 2013). Quanto aos ODVs, um complexo de proteínas de membrana denominadas fatores de infecção *per os* (PIF, do inglês '*per os* infective factor') são responsáveis pela fusão direta do envelope com a membrana das microvilosidades das células do epitélio do intestino do inseto hospedeiro (Slack & Arif, 2007). Esta fusão culmina na liberação de nucleocapsídeos no citoplasma celular. Importante, ODVs de *Alphabaculovirus* podem confinar um ou múltiplos nucleocapsídeos e são, por isso, respectivamente denominados SNPV (do inglês, 'single NPV') ou MNPV (do inglês, 'multiple NPV) (Rohrmann, 2014). O ganho evolutivo e os fatores moleculares que geram tais fenótipos não são claros; entretanto, já se é sabido que em MNPVs, após fusão do ODV com a microvilosidade, um nucleocapsídeo pode estabelecer a infecção na célula colunar e os outros podem sofrer transcitose e atravessar a célula para iniciar a infecção secundária (Rohrmann, 2014).

A construção da história evolutiva da família *Baculoviridae* permeia o estudo sistemático do vírus quanto à sua caracterização estrutural, patologia do inseto e da célula hospedeira bem como genômica e proteômica do vírus. De fato, a base para o início de quaisquer estudos moleculares mais detalhados de novas espécies virais ou de isolados certamente se inicia com sequenciamento do genoma completo. Assim, com o avanço das técnicas de sequenciamento de alto desempenho, novos genomas de baculovírus surgem de forma crescente permitindo um entendimento mais profícuo da história evolutiva da família viral. Além disso, é importante salientar que os dados gerados com sequenciamento influenciam diretamente no uso de baculovírus como agentes de controle biológicos bem como em seu melhoramento como vetor de expressão heteróloga. Por exemplo, análise da estabilidade genética de isolados temporais ou mutações associadas à perda ou ganho de virulência são informações

obtidas com a genômica de baculovírus que contribuem para o uso do vírus como controlador biológico. Além disso, a descoberta e caracterização de genes relacionados a desempenho viral pode aperfeiçoar a produção de proteínas heterólogas. Atualmente, existem mais de 100 genomas de baculovirus sequenciados e disponíveis no Genbank. Entretanto, apenas pouco mais de 60 são de espécies inéditas. Até o início deste trabalho (03/2012) existiam somente dois genomas de baculovírus isolados no Brasil sequenciados e publicados: o baculovirus da espécie *Anticarsia gemmatalis multiple nucleopolyhedrovirus* (AgMNPV) (Oliveira *et al.*, 2006) e o isolado brasileiro 19 da espécie *Spodoptera frugiperda multiple nucleopolyhedrovirus* (SfMNPV) (Wolff *et al.*, 2008).

2. Objetivos gerais

A fim de contribuir com o conhecimento mais amplo da diversidade viral de microrganismos isolados no Brasil, este trabalho teve por objetivo sequenciar e caracterizar novas espécies ou isolados de baculovírus brasileiros em níveis patológico, molecular, filogenético e estrutural.

3. Objetivos específicos

- Sequenciar e descrever o genoma do primeiro isolado não-asiático da espécie *Bombyx mori nucleopolyhedrovirus* (BmNPV), caracterizar estruturalmente o vírus, analisar a história evolutiva de genes *bro* (local de maior divergência entre os isolados) e a diversidade genética da população viral isolada (Capítulo 2).

- Sequenciar e descrever o genoma do betabaculovírus da espécie *Erinnyis ello granulovirus* (ErelGV), caracterizar estruturalmente e analisar a filogenia do vírus e de alguns genes adquiridos por transferência horizontal ou duplicação (Capítulo 3).

- Identificar um baculovírus isolado de *Helicoverpa armigera* durante o primeiro surto nacional da praga, sequenciar e descrever o genoma completo, caracterizar estruturalmente e identificar a diversidade nucleotídica da população sequenciada (Capítulo 4). Além disso, comparar a patogenia do vírus a uma cepa comercial.

- Caracterizar funcional e filogeneticamente um *inibidor de serino protease* (do inglês, *serpin*, '*serine protease inhibitor*') identificado no baculovírus da espécie *Hemileuca species nucleopolyhedrovirus* (HespNPV) (Capítulo 5 – projeto principal do Doutorado Sanduíche).

- Sequenciar e descrever o genoma do betabaculovírus da espécie *Diatraea saccharalis granulovirus* (DisaGV), caracterizar estruturalmente e analisar a filogenia do vírus e de alguns genes adquiridos por transferência horizontal como uma proteína GPCR vinda de inseto e um proteína de fusão de envelope nunca

observada em betabaculovírus (Capítulo 6). Além disso, caracterizar funcionalmente a proteína de fusão de envelope, gp64, encontrada no genoma de DisaGV (Capítulo 7).

- Sequenciar e descrever o genoma do baculovírus *Perigonia lusca single nucleopolyhedrovirus* (PeluSNPV), caracterizar estruturalmente o vírus, estabelecer filogenia e analisar a história evolutiva de um gene especial de metabolismo de nucleotídeo encontrado. Além disso, entender o papel deste gene na infecção viral e analisar sua funcionalidade (Capítulo 8).

4. Referência

Braconi, C. T., Ardisson-Araujo, D. M., Leme, A. F., Oliveira, J. V., Pauletti, B. A., Garcia-Maruniak, A., Ribeiro, B. M., Maruniak, J. E. & Zanotto, P. M. (2014). Proteomic analyses of baculovirus Anticarsia gemmatalis multiple nucleopolyhedrovirus budded and occluded virus and associated host cell proteins. *J Gen Virol*.

Clem, R. J. & Passarelli, A. L. (2013). Baculoviruses: sophisticated pathogens of insects. *PLoS Pathog* **9**, e1003729.

Herniou, E. A. & Jehle, J. A. (2007). Baculovirus phylogeny and evolution. *Curr Drug Targets* **8**, 1043-1050.

Jehle, J. A., Blissard, G. W., Bonning, B. C., Cory, J. S., Herniou, E. A., Rohrmann, G. F., Theilmann, D. A., Thiem, S. M. & Vlak, J. M. (2006). On the classification and nomenclature of baculoviruses: a proposal for revision. *Arch Virol* **151**, 1257-1266.

Oliveira, J. V., Wolff, J. L., Garcia-Maruniak, A., Ribeiro, B. M., de Castro, M. E., de Souza, M. L., Moscardi, F., Maruniak, J. E. & Zanotto, P. M. (2006). Genome of the most widely used viral biopesticide: Anticarsia gemmatalis multiple nucleopolyhedrovirus. *J Gen Virol* **87**, 3233-3250.

Ooi, B. G. & Miller, L. K. (1988). Regulation of host RNA levels during baculovirus infection. *Virology* **166**, 515-523.

Ribeiro, B. M.; Morgado, Fabrício da Silva; Ardisson-Araújo, Daniel Mendes Pereira; Silva, Leonardo A.; Cruz, F. S. P.; Chaves, L. S. C.; Quirino, M. S.; Andrade, M. S.; Corrêa, R. F. T.. Baculovírus para expressão de proteínas recombinantes em célula de inseto. In: Rodrigo R. Resende. (Org.). Biotecnologia Aplicada à Saúde. 1ed.São Paulo: Blucher, 2015, v. 2, p. 252-306.

Rohrmann, G. F. (2013). *Baculovirus Molecular Biology*. Bethesda (MD): National Center for Biotechnology Information (US). Available from: http://www.ncbi.nlm.nih.gov/books/NBK49500/.

Rohrmann, G. F. (2014). Baculovirus nucleocapsid aggregation (MNPV vs SNPV): an evolutionary strategy, or a product of replication conditions? *Virus Genes*.

Slack, J. & Arif, B. M. (2007). The baculoviruses occlusion-derived virus: virion structure and function. *Adv Virus Res* **69**, 99-165.

Summers, M. D. (2006). Milestones leading to the genetic engineering of baculoviruses as expression vector systems and viral pesticides. *Adv Virus Res* **68**, 3-73.

Wang, M., Wang, J., Yin, F., Tan, Y., Deng, F., Chen, X., Jehle, J. A., Vlak, J. M., Hu, Z. & Wang, H. (2014). Unraveling the entry mechanism of baculoviruses and its evolutionary implications. *J Virol* **88**, 2301-2311.

Wolff, J. L., Valicente, F. H., Martins, R., Oliveira, J. V. & Zanotto, P. M. (2008). Analysis of the genome of Spodoptera frugiperda nucleopolyhedrovirus (SfMNPV-19) and of the high genomic heterogeneity in group II nucleopolyhedroviruses. *J Gen Virol* **89**, 1202-1211.

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

Capítulo 2. Complete genome sequence of the first non-Asian isolate of *Bombyx mori nucleopolyhedrovirus*

1. Abstract

Brazil is one of the largest silk producers in the world. The domesticated silkworm (*Bombyx mori*) was formally introduced into the country in the twentieth century and the state of Paraná is the main national producer. During larval stages, *B. mori* can be afflicted by many different infectious diseases, which lead to substantial losses in silk production. In this work, we describe the structure and complete genome sequence of the first non-Asian isolate of *Bombyx mori nucleopolyhedrovirus* (BmNPV), the most important silkworm pathogen. The BmNPV-Brazilian isolate is a nucleopolyhedrovirus with singly enveloped nucleocapsids within polyhedral occlusion bodies. Its genome has 126,861 bp with a G+C content of 40.4%. Phylogenetic analysis clustered the virus with the Japanese strain (BmNPV-T3). As expected, we have detected intra-population variability in the virus sample. Variation along homologous regions (HRs) and *bro* genes was observed; there were seven HRs, deletion of *bro-e*, and division of *bro-a* into two ORFs. The study of baculoviruses allows for a better understanding of virus evolution providing insight for biological control of insect pests or protection against the pernicious disease caused by these viruses.

Key-words: *Bombyx mori*; complete genome; baculovirus; BmNPV isolate; intra-isolate diversity.

Este capítulo foi publicado na revista Virus Genes. **Ardisson-Araujo, D. M., Melo, F. L., de Souza Andrade, M., Brancalhao, R. M., Bao, S. N. & Ribeiro, B. M. (2014).** Complete genome sequence of the first non-Asian isolate of Bombyx mori nucleopolyhedrovirus. *Virus Genes* **49**, 477-484.

2. Introduction

The *Baculoviridae* is a diverse family of insect viruses with circular double-stranded genomic DNA (Rohrmann, 2013). They are divided phylogenetically into four genera: *Alpha*, *Beta*, *Gamma* and *Deltabaculovirus* (Jehle *et al.*, 2006). Both *Alpha* and *Betabaculovirus* produce occlusion-derived virions (ODVs) and budded virions (BVs) during a complete infection cycle (Slack & Arif, 2007). ODVs are orally infectious and are protected within a crystalline protein matrix called occlusion body (OB). After ingestion of contaminated food by the larvae, the OB dissolution releases ODVs that infect primarily the insect midgut epithelia (Xu *et al.*, 2010). BVs are produced early in the replicative cycle (Wang *et al.*, 2010) and disseminate from the midgut to the entire insect body (Washburn *et al.*, 2002). In the end of infection, the larvae die and release OBs to the environment. The environmental stability of ODVs in OBs, the host specificity, and the lethality of infection make baculoviruses important pathogens for both beneficial and pest insects (Summers, 2006; Vasyl'ieva & Lebedynets, 2001).

Almost five hundred alphabaculovirus have been described (Jehle *et al.*, 2006) and the genomes of more than sixty have been fully sequenced (Rohrmann, 2013). Among these genomic data, there are ten Asian isolates found to infect the genus *Bombyx* L. 1758 (Lepidoptera: Bombycidae). Two were isolated from *B. mandarina* (BomaNPV-S1 and

-S2) (Cheng *et al.*, 2012; Xu *et al.*, 2010) the silkworm found in nature, and eight were isolated from the domesticated silk thread producer, *B. mori* (BmNPV-T3, -Cubic, -Indian, -Zhejiang, -Guangxi, -C1, -C2, and -C6) (Cheng *et al.*, 2012; Fan *et al.*, 2012; Gomi *et al.*, 1999; Xu *et al.*, 2013).

B. mori is able to weave a big cocoon for protection during metamorphosis (Pandiarajan *et al.*, 2011). This structure is composed of a single thread and can be used for fabric manufacture (Blossman-Myer & Burggren, 2009). Human intervention directed the insect evolution by inbreeding and artificial selection in order to increase silk production (Doreswamy & Gopal, 2013). As a result, the imago became unable to fly, mate, or even feed by itself. In other words, the domesticated silkworm is completely dependent on humans for survival and has therefore become part of human culture (Ball, 2009). The history of silk is not restricted to the Asia. Brazil in South America is one of the largest commercial silk producers in the world. In 2009, almost five tons of cocoons were produced, according to the EMATER (Brazilian Government Company of Technical Assistance and Rural Extension). Interestingly, fourteen different strains of *B. mori* have been identified in Brazil and biological assays have demonstrated that a Brazilian BmNPV (called here BmNPV-Brazilian) was found infecting these different commercial strains (Brancalhao *et al.*, 2009). BmNPV is the major cause of silk production losses and is a serious problem for sericulture in Brazil and in all other silk-producing countries (Brancalhao *et al.*, 2009; Pereira *et al.*, 2013). Therefore, in order to better understand this important pathogen, we describe here the complete genome sequence of the BmNPV-Brazilian.

3. Materials and Methods

3.1. Insect infection

Fourth *instar B. mori* hybrid caterpillars were obtained from the silk industry (Fiação de Seda BRATAC S.A., Paraná, Brazil) and raised on fresh mulberry leaves (*Morus sp*.) as previously described (Pereira *et al.*, 2008). BmNPV was obtained from infected *B. mori* hybrid caterpillars found in Paraná state in Brazil (Brancalh,,o, 2002). Fifth instar larvae were starved for 24 hours after ecdysis and fed on mulberry leaf discs (2 cm diameter) with 20 µl of viral suspension at a concentration of $8x10^8$ OBs/ml for virus amplification as previously described (Ribeiro Lde *et al.*, 2009). Following complete ingestion, caterpillars were placed in individual plastic cups.

3.2. Virus purification, Bm-5 cell infection, and DNA extraction

Insect cadavers were collected and homogenized with the same volume of ddH_2O (w/v), filtered through three layers of gauze, and centrifuged at 7,000 *x g* for 10 min. The pellet was washed three times with SDS 0.5% (w/v) and once with NaCl 0.5 M followed by centrifugation at 7,000 x g for 10 min for each washing. The last washed-resulting pellet was resuspended in ddH_2O, loaded onto a continuous 20-65% sucrose gradient, and centrifuged at 104.000 *x g* for three hours at 4 ºC. The OB band was collected, 3-fold diluted in ddH_2O, and centrifuged at 7,000 *x g* for 15 min at 4 ºC. Purified polyhedra (10^9 OBs/ml) were dissolved in an alkaline solution and used for both Bm-5 cell monolayer infection and to extract DNA. Bm-5 cells were maintained at 28 ºC in TNMFH (GIBCO BRL Life Technologies), supplemented with 10% fetal

bovine serum (Invitrogen, Carlsbad, CA, USA). DNA was extracted according to O'Reilly *et al*. (O'Reilly *et al.*, 1992) from ODVs. The quantity and quality of the isolated DNA were determined by electrophoresis on a 0.8% agarose gel (data not shown).

3.3. Ultrastructural analyses

For Scanning Electron Microscopy (SEM), OBs (10^9 OBs/ml) were treated with acetone 1 X and then incubated at 25 ºC for 1 hour. The solution was loaded in a metallic stub, dried overnight at 37 ºC, coated with gold in a Sputter Coater (Balzers) for 3 min, and observed in a SEM Jeol JSM 840A at 10 kV. For Transmission Electron Microscopy (TEM), pellets of purified OBs were fixed in Karnovsky fixative (2.5% glutaraldehyde, 2% paraformaldehyde, in 0.1 M, pH 7.2, cacodylate buffer) for 2 h, post-fixed in 1% osmium tetroxide in the same buffer for 1 h and then stained *en bloc* with 0.5% aqueous uranyl acetate, dehydrated in acetone, and embedded in Spurr's low viscosity embedding medium. The ultrathin sections were contrasted with uranyl acetate/lead citrate and observed in a TEM Jeol 1011 at 80 kV.

3.4. Genome sequencing, annotation and analysis

BmNPV-Brazilian (hereafter designated as Brazilian) genomic DNA was sequenced with the 454 Genome Sequencer (GS) FLX™ Standard (Roche) at Macrogen (Seoul, Korea). The singe-end reads were analyzed using Geneious 6.0 (Kearse *et al.*, 2012). Firstly, all the reads were trimmed to remove sequencing adaptor and low quality regions (Q≥20), and then assembled *de novo* using a minimum overlap parameter of 200

nt and minimum overlap identity of 98%. The resulting contigs were mapped on the genome of BmNPV-T3 isolate (hereafter designated as T3) (Table 1) (Cheng *et al.*, 2012; Fan *et al.*, 2012; Gomi *et al.*, 1999; Xu *et al.*, 2013; Xu *et al.*, 2010). Next, the consensus sequence was used in a reference-guided alignment to obtain the consensus genome of our isolate. The Genbank accession number is KJ186100. Frame shifts at homopolymeric regions introduced by the 454-pyrosequencing method were corrected manually. For genome annotation, only open reading frames (ORFs) with at least 150 nucleotides (nt) were considered. The homologous proteins were identified using blastp (Altschul & Lipman, 1990). For phylogenetic analysis, a MAFFT alignment (Katoh *et al.*, 2002) was carried out with whole genome sequences of all *Bombyx*-isolate baculoviruses available in Genbank (Table 1) and the AcMNPV-C6 genome (L22858.1). This alignment was manually inspected and poor aligned regions(at least 50% of gaps) were deleted. The resulting alignment was approximately 127 kb long. Maximum likelihood tree was inferred using RAxML(Stamatakis *et al.*, 2008) and PhyML (Guindon *et al.*, 2010), under the Tamura-Nei model selected by jModelTest-2.1.4 (Darriba *et al.*, 2012). The branch support was estimated by non parametric bootstrap analysis with 100 repetitions (Stamatakis *et al.*, 2008) and Shimodaira-Hasegawa-like test (Anisimova *et al.*, 2011). Moreover, a gene comparison was performed using all *Bombyx*-isolate baculovirus (Table 1). This dataset was compared using CGView Comparison Tool (Ardisson-Araujo *et al.*, 2013) and the results were plotted using CIRCOS. Moreover, single nucleotide polymorphisms (SNP) were detected using Geneious 6.0. To perform this analysis the trimmed reads were mapped to the Brazilian isolate genome and the SNP were identified using the following parametrs: *p*-value for the sequence error of $1x10^{-6}$, a minimum coverage of 20 reads,

and a minimum variant frequency of 0.25. The lower the *p*-value is, the more likely the SNP represents an authentic variation.

4. Results

4.1. Ultrastructural analyses and *B. mori*-derived cell infection

In this work we described the first non-Asian isolate of the baculovirus species BmNPV. The baculovirus was infecting a strain of the silkworm *B. mori* reared in Brazil for silk industry (Fig. 1a). OBs were purified from larvae cadavers and used for ultrastructural analyses. We observed single-occluded virions inside the protein matrix by Transmission Electron Microscopy (TEM) (Fig. 1c) and polyhedral OB shape by Scanning Electron Microscopy (SEM) (Fig. 1f). In general, the OBs presented size of 2 to 4 µm with a regular shape. Immature OBs were also observed among the sample with spaces for ODV occlusion (Fig. 1f, inset). BmNPV is infectious to *B. mori*-derived cells such as the strain Bm-5 (Grace, 1967). Therefore, we used ODVs released from alkaline solution-treated OBs to infect Bm-5 cells. Infected cells presented typical features of baculovirus infection (Rohrmann, 2013) with nuclear hypertrophy and cell rounding (data not shown) and at late time post-infection several polyhedra were observed inside the cell nucleus (Fig. 1b). Interestingly, as previously described (Brancalh,,o, 2002), we also observed ODVs with multiple nucleocapsids (Fig. 1d) and few irregular-shaped polyhedra (Fig. 1e).

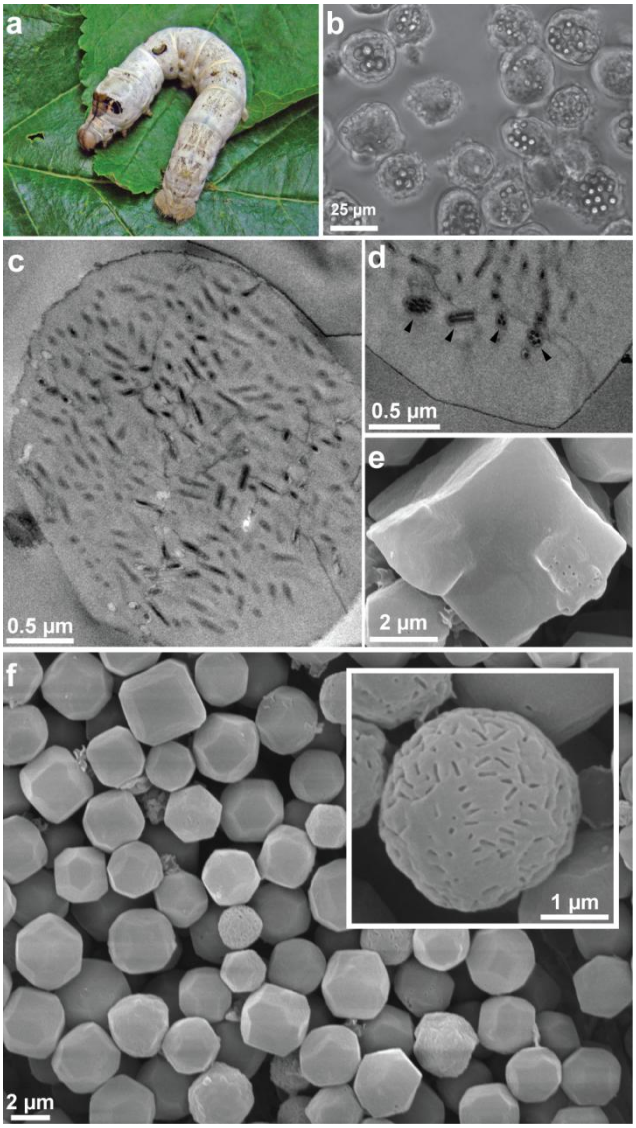

Fig. 1 Silkworm strain reared in Brazil, cell infection, and ultrastructural analysis of occlusion bodies (OBs) from BmNPV-Brazilian isolate. **a** A silkworm reared in the Brazilian silk industry feeding on a mulberry leaf. **b** Bm-5 cells infected with the Brazilian isolate at 72 h p.i.. **c** Transmission electron micrograph reveals OB with single nucleocapsids (nc) within. **d** Polyhedra containing both single and multiple embedded rod-shaped nucleocapsids within single ODVs (arrowhead). **e** Scanning electron micrograph of a tetrahedral OB observed in our sample. **f** Several OBs with polyhedral shape and an inset showing immature OBs with holes for ODV occlusion.

4.2. Genome features, phylogenetic analysis, and gene comparison

The 454 sequencing produced approximately 23,000 single-end reads. After size and quality trimming, 18,240 reads (average size of 496 nt) were assembled *de novo* with coverage of 46.7 ± 12.9. The genome has a size of 126,861 bp and a G+C content of 40.4%, which is close to the average size of 127,159 ± 1,158.5 bp for *Bombyx*-isolated viruses (Table 1). The pairwise identity between BmNPV-Brazilian and the remaining isolates varied from 97.9 to 96.3 (Table 1). Our phylogenetic analysis shows that both BmNPV isolates and BomaNPV-S1 form together a well-supported monophyletic clade (Fig. 2), as previously described (Xu *et al.*, 2010). Moreover, the newly sequenced Brazilian isolate clustered with BmNPV-T3 strain, originally isolated from Japan (Fig. 2). This found is compatible with a virus introduction from Japan to Brazil. Annotation of the Brazilian genome resulted in 143 ORFs with more than 150 nt. As shown in Figure 3 and Table S1, most of these ORFs are shared among the *Bombyx*-isolated viruses, as well as with AcMNPV. The only unique ORF was the *Bm(Br)Orf-26*, which encodes a putative protein of 80 amino acids with no homologous in GenBank. Most variations were due to deletions and insertions on homologous repeat regions (hr) (Fig. 3, green color and Table S2) and *baculovirus repeated orf* (*bro*) genes (Fig. 3, red color). Seven HRs were identified in the BmNPV-Brazilian genome. We found a large deletion in HR2L compared to the other isolates. HR2L and HR2R flank both the *Bm(Br)Orf-26* and the *fgf* gene (Fig. 3). Even presenting high number of insertions and deletions, the identity among the HRs remained high among the isolates during pairwise alignment analyses (Table S2). The lowest global identity was observed for HR4L with 74.9% of identity and an average size of 361.6 ± 75.5 bp. Both Guangxi and Zheijang isolates presented a complete synapomorphic deletion of the HR2L. On the other hand,

the closest relative to the Brazilian isolate (isolate T3) presented two insertions and no deletion at that same HR. Regarding the *bro* gene variability, a notable aspect was a division of *bro-a* into two ORFs (*bro-a1*: *Bm(Br)Orf-23* and *bro-a2*: *Orf-24*) (Fig. 3, in red color), due to a single nucleotide polymorphism that introduced a stop codon (TAG). To confirm this, we searched carefully among the reads and identified 59 out of 75 reads presenting the stop codon-introducing polymorphism (TCG to TAG) into the *bro-a* coding region, suggesting it as an authentic polymorphism.

Table 1. *Bombyx*-isolated genomes used in this study

Virus-Strain	Size (nt)	Id (%)[a]	Country	Reference	Accession number
BmNPV isolates					
Brazilian	126,863	100	Brazil	This work	KJ186100
T3	128,413	97.2	Japan	[16]	L33180.1
Guangxi	126,843	97.9	China	[20]	JQ991011
Zheijiang	126,125	97.6	China	[20]	JQ991008
C1	127,901	96.3	South Korea	u/d	KF306215
C2	126,406	97	South Korea	u/d	KF306216
C6	125,437	96.6	South Korea	u/d	KF306217
Cubic	127,465	96.3	China	[19]	JQ991009
India	126,879	97.9	India	[21]	JQ991010
BomaNPV isolates					
S1	126,770	97.9	China	[22]	FJ882854.1
S2	129,646	93.7	China	[23]	JQ071499.1

u/d - unpublished data.
a - identity related to whole genome of BmNPV-Brazilian isolate by MAFFT Alignment (25)

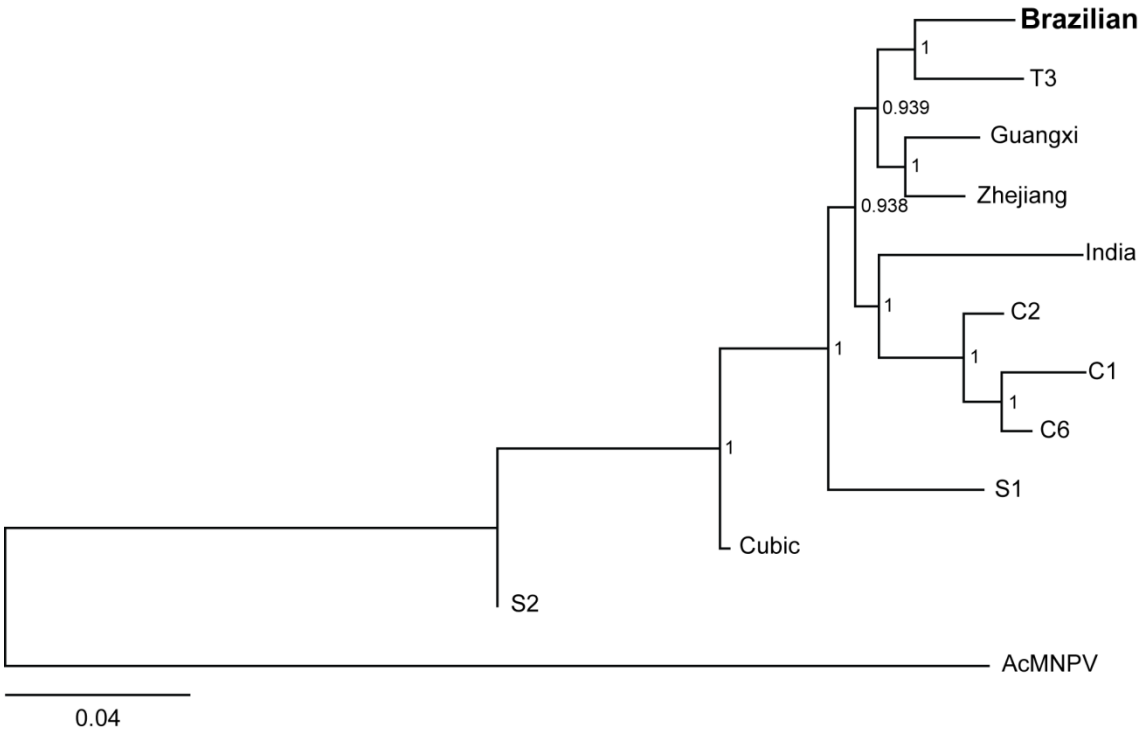

Fig. 2 Maximum likelihood tree for *Bombyx*-isolated baculoviruses. The phylogenetic inference is based on the MAFFT alignment among the whole genome using PhyML method (27). The AcMNPV is used as outgroup. The Brazilian (in bold) isolate is closely related to the Japanese plaque-isolated virus, T3. The branch support is estimated by a Shimodaira-Hasegawa-like test (29).

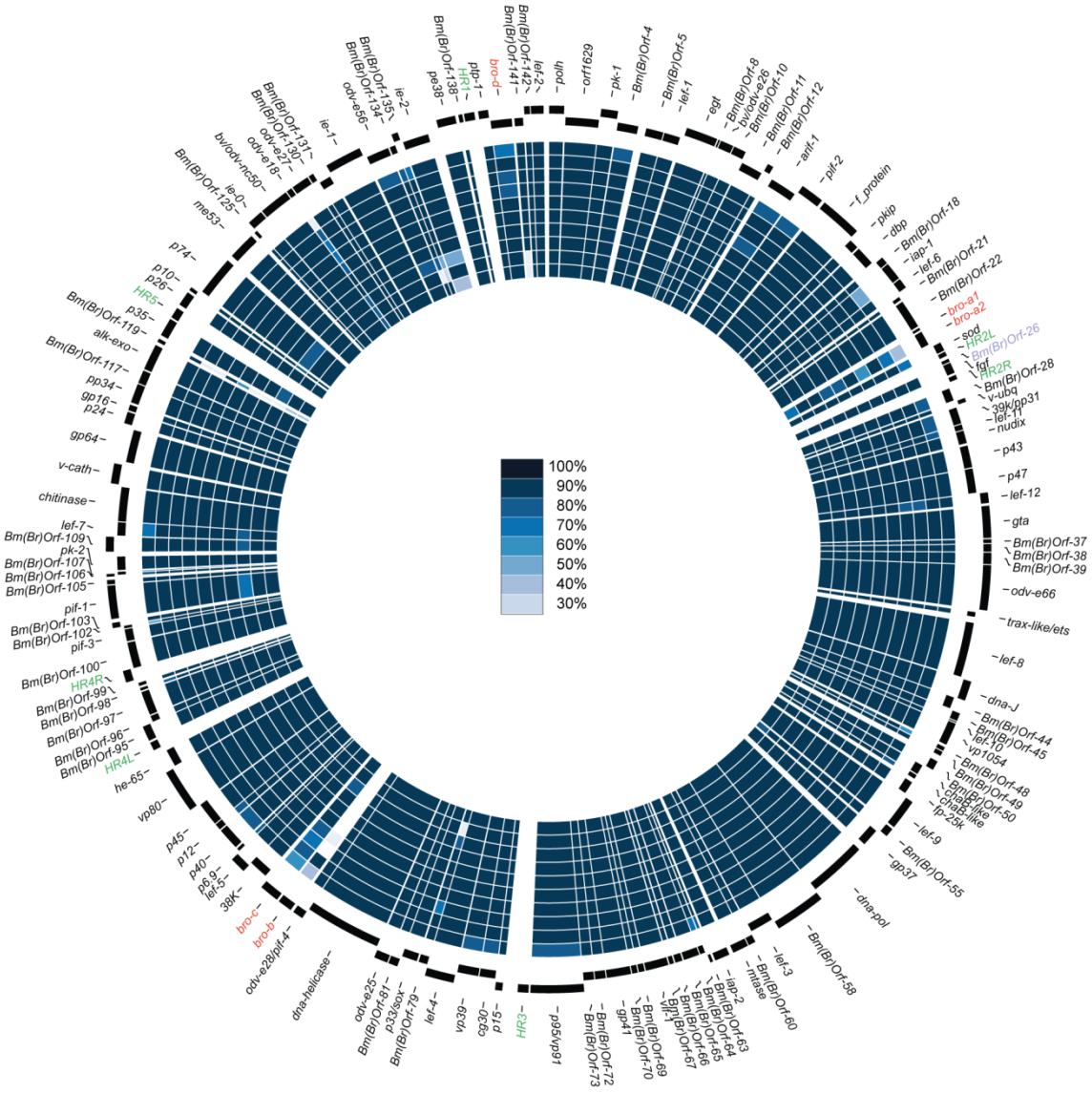

Fig. 3 Gene comparison of the BmNPV-Brazilian isolate with all the *Bombyx*-isolated baculoviruses. The heat map shows a comparison between both all CDS from Brazilian isolate and from all *Bombyx*-isolated baculoviruses. The identity (from 0 to 100%) is plotted in shades of blue for the ten inner circles. From the outermost ring: T3, Guangxi, Cubic, S1, India, C1, S2, C2, C6, Zhejiang. HRs are plotted in green only for the Brazilian isolate, *bro* genes are in red, and the *Bm(Br)Orf-26* (without homologues in NCBI) is highlighted in purple. Only the isolates Brazilian, T3, and Guangxi present the *bro-b* gene.

4.3. The gain and loss of *bro* genes

The distribution of *bro* genes along the phylogeny and a gene context analysis are shown in Figure 4. Four major observations can be draw from this distribution: (*i*) the most recent common ancestor (MRCA) of AcMNPV and *Bombyx*-isolated viruses had the *bro-d* gene (Fig. 4a and Fig. 4b);(*ii*) *bro-a*, *bro-c*, and *bro-e* were probably gained by the MRCA of all *Bombyx*-isolated viruses (Fig. 4a); (*iii*) *bro-e* was lost in several isolates (Fig. 4a and 4b); (*iv*) *bro-b* was gained by the MRCA of the isolates Brazilian, T3, Guangxi and Zheijiang, and was subsequently lost by Zheijiang (Fig. 4a and 4b). Thus, *bro-a*, *bro-c* and *bro-d* were conserved in all *Bombyx*-isolated viruses, except for *bro-a* in both C6 (partial deletion) and Brazilian (split in two ORFs, as described above) isolates. The *bro-b* and *bro-e* were present only in a small number of isolates. Such pattern of gene evolution is compatible with multiple events of gene duplication and losses, as previously suggested by Kang et al. (Kang *et al.*, 1999). A phylogenetic analysis using an alignment of all predicted BRO proteins confirmed that *bro-a* and *bro-c* are closely related as well as *bro-b* and *bro-e* (Fig. S1). Therefore, it is reasonable to assume that *bro-b* originated probably from a *bro-e* duplication event in the ancestral lineage of the isolates Brazilian, T3, Guangxi and Zheijiang. Conversely, the *bro-e* evolutionary history was probably the result of one ancestral gain followed by six independent losses in several isolates (Fig. 4a). This independent loss scenario is corroborated by the gene context analysis, which showed that all isolates with *bro-e*, complete or vestigial, presented the same genomic context (Fig. 4b).

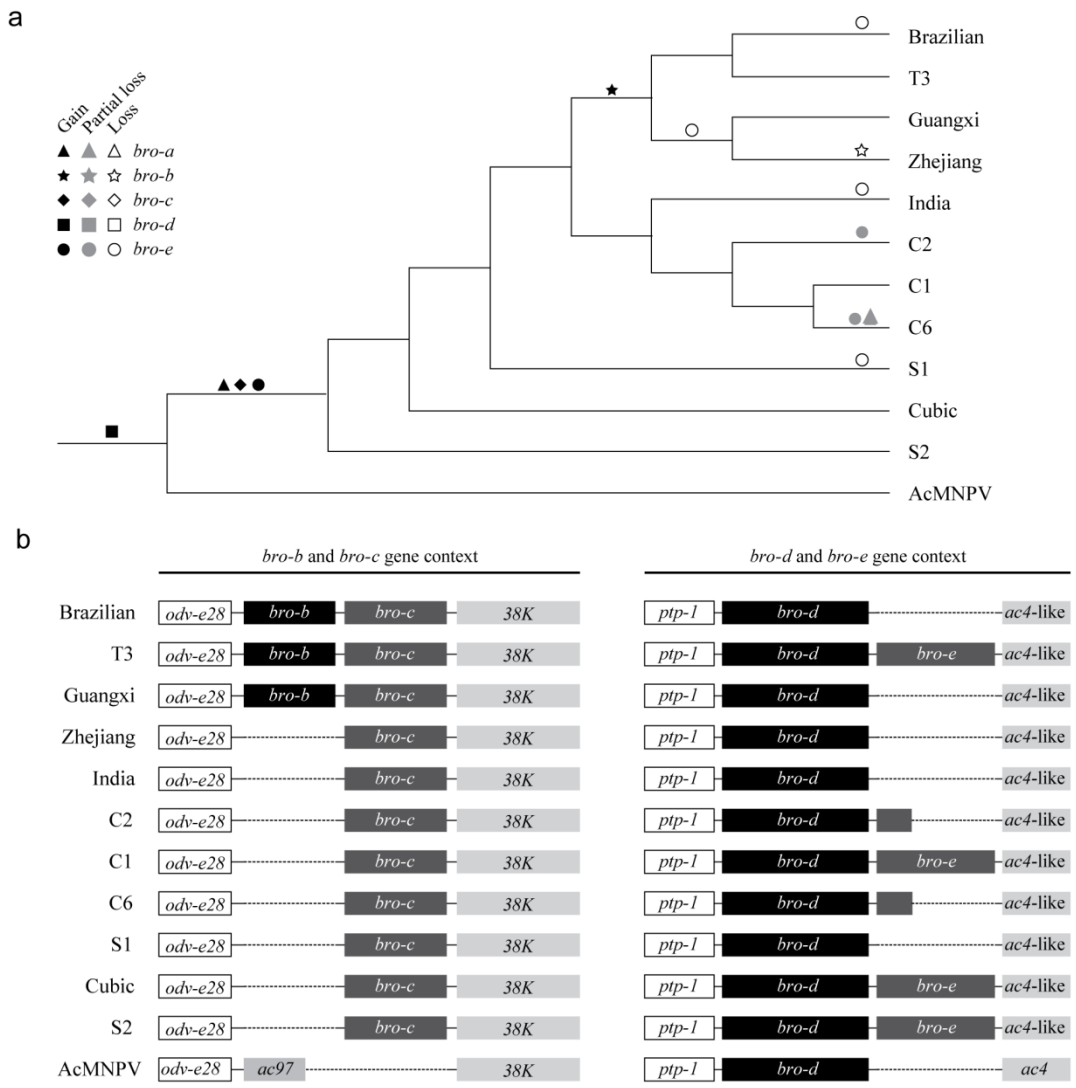

Fig. 4 *bro* genes occurrence among *Bombyx*-isolated baculoviruses. **a** Gain, partial loss, and loss (black, gray and empty symbols, respectively) for *bro* genes along the evolutionary history of *Bombyx*-isolated baculoviruses. AcMNPV is used as outgroup. Brazilian isolate presents a division of the *bro-a* gene into two orfs (not shown) and the C6 isolate presents a partial deletion (gray triangle). **b** Gene context of the *bro-b*, *bro-c*, *bro-d*, and *bro-e* genes.

4.4. Intra-isolate diversity in BmNPV-Brazilian

As previously described, the 454 sequencing of the Brazilian isolate resulted in 18,240 reads that were used to assembly the complete genome. However, this reads also provided information at genotypic variation within the isolate. Although the sequencing coverage of 46.7 ± 12.9, we were able to identify 404 SNPs, ignoring insertions and deletions associated frequently with 454-pyrosequencing errors. As shown in Fig. S2, most polymorphism observed was synonymous (67%). It was possible to observe a high number of SNPs in *Bm(Br)Orf-74* (*p95* with 23 SNPs), *Bm(Br)Orf-15* (*f protein*, with 13 SNPs), *Bm(Br)Orf-83* (*dna-helicase*, with 11 SNPs), and *Bm(Br)Orf-71* (*gp41*, with 9 SNPs).

5. Discussion

In this study, the genome of a BmNPV strain isolated in Brazil was sequenced and compared to distinct *Bombyx*-isolated baculoviruses. The Brazilian strain is closely related to the strain T3, a Japanese isolate. The sericulture introduction history in Brazil is not clear. Some documents point to Japanese immigrants as the first formal silk producers in the country. Here, our find has suggested that both caterpillar and virus could have been introduced from Japan to Brazil.

Most of genomic divergences among BmNPV isolates were in HRs and also in *bro* genes. Both regions were previously identified as primary areas of divergence within genomes of *Bombyx*-isolated baculoviruses (Xu *et al.*, 2013). Here, we observed HR size variance in the genome of BmNPV-Brazilian. HRs are imperfect palindromic

sequences with size and location highly variable. The Brazilian isolate presented seven HRs with a large deletion in the HR2L which has been identified as the most unstable HR in BmNPV genomes (Xu *et al.*, 2013). HRs are believed to play roles in genome replication, recombination, and gene transcription (Rohrmann, 2013).

Sequence plasticity is also true for *bro* genes, being a constant trait for baculovirus genome evolution. These features are observed among different baculovirus species such as in virus isolated from *Helicoverpa armigera* and *Spodoptera frugiperda* (Harrison & Popham, 2008; Ogembo *et al.*, 2009; Rowley *et al.*, 2011; Simon *et al.*, 2011; Zhang *et al.*, 2013). In this work, we found that the *bro-a* gene is divided into two ORFs. Previous work found insertions and deletions inside *bro-a*, during gene comparison between BmNPV-T3 and other plaque-purified BmNPV isolates (Pang *et al.*, 2007). Moreover, the isolate BmNPV-C6 presents a partial deletion at the carboxi-terminal of the *bro-a* gene, suggesting that this region is probably not required for virus viability. In the specific case of *Bombyx*-isolated baculoviruses, genome insertions and deletions (indels) of *bro* genes are quite common (Kang *et al.*, 1999) (Fig. 3). These indels have been implicated in viral pathogenicity, genome replication capacity, and/or viral gene transcription kinetics (Xu *et al.*, 2013; Zemskov *et al.*, 2000). Since *bro* genes present a high repetitive content, the phylogenetic reconstruction can be misinterpreted based only in the gene sequence; hence we also looked at the loci of the genes. The AcMNPV genome is closest to *Bombyx*-isolated baculoviruses and present only one *bro* gene, a homologous to *bro-d* (Fig. 4b). In fact, the *Bombyx*-isolated baculovirus *bro* genes could be result of several duplication events that occurred only after the ancestral split of these lineages. The *bro* genes belong to a unique multigenic family (Bideshi *et al.*, 2003). AcMNPV, as explained above, contains only a single *bro* gene in its genome

(Ayres *et al.*, 1994). On the other hand, the *Spodoptera exigua multiple nucleopolyhedrovirus* (SeMNPV) completely lacks *bro* genes (Wf *et al.*, 1999) and in *Bombyx*-isolated viruses the amount of *bro* genes varies (Fig. 4). Interestingly, some BRO proteins are present in both the cytoplasm and nuclei of infected cells (Gong *et al.*, 2003; Kang *et al.*, 1999). They have nucleic acid and nucleosome association capabilities (Zemskov *et al.*, 2000), a single-stranded DNA (ssDNA) binding motif (Zemskov *et al.*, 2000), and can also be present or not as a component of the virion structure (Braconi *et al.*, 2014; Deng *et al.*, 2007; Gong *et al.*, 2003; Perera *et al.*, 2007; Wang *et al.*, 2010; Xu *et al.*, 2013). However, the specific functions of *bro* genes and their protein products are still unknown. Specific *bro* genes seems to be crucial for the virus replication, considering the evolution with its own host, such as *bro-d* and *bro-c* genes of BmNPV (Kang *et al.*, 1999) conserved in all isolates (Fig. 4A).

We also found genomic diversity in our sample. The genotypic variation among viruses isolated from the field, in this situation, from the silk industry, is a common feature of baculoviruses (Craveiro *et al.*, 2013). We did not plaque-purify the virus in order to access its diversity. Plaque-isolated viruses do not reflect intra-population heterogeneity and may also introduce errors or privilege genotypes during *in vitro* cell replication, changing drastically the virus diversity or introducing new errors. In fact, the 454 sequencing may cause errors reflecting in a false variability. Therefore, we consider variation base on a minimum coverage of 20-fold, a minimum variant frequency of 0.25 with a *p*-value for the sequence error of $1x10^{-6}$, meaning that the chance to see a variant by chance is 0.0001%. Intra-specific diversity might somehow be reflected in phenotypic features, for example the capacity of a single nucleopolyhedrovirus, as BmNPV isolates are, to occlude more than one nucleocapsid per virion (Fig. 1b, in the

same OB is possible to see multiple and single ODVs, showing this is not a contamination) or production of abnormal-shaped polyhedra (Fig. 1c). Interestingly, we found high number of SNPs in the genes *p95, dna-helicase*, and *gp41* which are core genes in the family *Baculoviridae* (Garavaglia *et al.*, 2012). However, the impact of this diversity in virus replication or pathogenicity is not clear. P95 has shown to be essential for BV production and nucleocapsid assembly (Xiang *et al.*, 2013) being an ODV-associated structural protein (Braunagel *et al.*, 2003) and a component of the *per os* infectivity factor (PIF) complex (Peng *et al.*, 2012). Moreover, DNA-helicase is an essential protein for virus replication (Ono *et al.*, 2012; Rohrmann, 2013) and GP41 is a tegument-associated glycoprotein important for BV production and virus spread efficiency (Ono *et al.*, 2012). Conversely, previous work has showed that different BmNPV isolates had a high degree of sequence divergence in ORFs, which are not core genes, but otherwise might play an important role in the virus evolution (Xu *et al.*, 2013). For instance, F protein, which is shared only among *Alpha* and *Betabaculovirus*, was found to present high level of SNPs as well (Garavaglia *et al.*, 2012). The protein is believed to be a non-essential remnant protein in BmNPV-like viruses (Group I *Alphabaculovirus*) playing a role only on the virus pathogenicity (Lung *et al.*, 2003). Therefore, the SNPs found might have influence in the adaptation of the virus to new strains of *B. mori* or other insect hosts.

6. Conclusion

The most informative way of accessing robust information about the evolutionary history of a virus is sequencing its whole genome. Overall, BmNPV is a good model for the study of baculovirus genome evolution since this virus is associated with an insect

that has been domesticated and reared by man for more than 3,000 years. Here, we described the first genome of a non-Asian isolate of the baculovirus species BmNPV.

7. Acknowledgements

We thank *Conselho Nacional de Desenvolvimento Científico e Tecnológico* (CNPq), *Fundação de Apoio à Pesquisa do Distrito Federal* (FAPDF), *Coordenação de Aperfeiçoamento de Pessoal de Nível Superior* (CAPES) for financial support; Ingrid Gracielle Martins da Silva for kindly helping with the sample for TEM and SEM; and Jeffrey J. Hodgson for kindly reviewing the English writing style.

8. Reference

Altschul, S. F. & Lipman, D. J. (1990). Protein database searches for multiple alignments. *Proc Natl Acad Sci U S A* **87**, 5509-5513.

Anisimova, M., Gil, M., Dufayard, J. F., Dessimoz, C. & Gascuel, O. (2011). Survey of branch support methods demonstrates accuracy, power, and robustness of fast likelihood-based approximation schemes. *Syst Biol* **60**, 685-699.

Ardisson-Araujo, D. M., Rocha, J. R., da Costa, M. H., Bocca, A. L., Dusi, A. N., de Oliveira Resende, R. & Ribeiro, B. M. (2013). A baculovirus-mediated strategy for full-length plant virus coat protein expression and purification. *Virol J* **10**, 262.

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

Ball, P. (2009). Rethinking silk's origins. *Nature* **457**, 945.

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

Blossman-Myer, B. & Burggren, W. W. (2009). The silk cocoon of the silkworm, Bombyx mori: macro structure and its influence on transmural diffusion of oxygen and water vapor. *Comp Biochem Physiol A Mol Integr Physiol* **155**, 259-263.

Braconi, C. T., Ardisson-Araujo, D. M., Leme, A. F., Oliveira, J. V., Pauletti, B. A., Garcia-Maruniak, A., Ribeiro, B. M., Maruniak, J. E. & Zanotto, P. M. (2014). Proteomic analyses of baculovirus Anticarsia gemmatalis multiple nucleopolyhedrovirus budded and occluded virus and associated host cell proteins. *J Gen Virol*.

Brancalh„o, R. M. C. (2002). VÌrus entomopatogÍnicos no bicho-da-seda: taxonomia e citopatologia causada por Nucleopolyhedrovirus em cÈlulas de Bombyx mori. *Biotechnology* **24**, 5.

Brancalhao, R. M., Torquato, E. F. & Fernandez, M. A. (2009). Cytopathology of Bombyx mori (Lepidoptera: Bombycidae) silk gland caused by multiple nucleopolyhedrovirus. *Genet Mol Res* **8**, 162-172.

Braunagel, S. C., Russell, W. K., Rosas-Acosta, G., Russell, D. H. & Summers, M. D. (2003). Determination of the protein composition of the occlusion-derived virus of Autographa californica nucleopolyhedrovirus. *Proc Natl Acad Sci U S A* **100**, 9797-9802.

Cheng, R. L., Xu, Y. P. & Zhang, C. X. (2012). Genome sequence of a Bombyx mori nucleopolyhedrovirus strain with cubic occlusion bodies. *J Virol* **86**, 10245.

Craveiro, S. R., Melo, F. L., Ribeiro, Z. M., Ribeiro, B. M., Bao, S. N., Inglis, P. W. & Castro, M. E. (2013). Pseudoplusia includens single nucleopolyhedrovirus: Genetic diversity, phylogeny and hypervariability of the pif-2 gene. *J Invertebr Pathol* **114**, 258-267.

Darriba, D., Taboada, G. L., Doallo, R. & Posada, D. (2012). jModelTest 2: more models, new heuristics and parallel computing. *Nat Methods* **9**, 772.

Deng, F., Wang, R., Fang, M., Jiang, Y., Xu, X., Wang, H., Chen, X., Arif, B. M., Guo, L. & Hu, Z. (2007). Proteomics analysis of Helicoverpa armigera single nucleocapsid nucleopolyhedrovirus identified two new occlusion-derived virus-associated proteins, HA44 and HA100. *J Virol* **81**, 9377-9385.

Doreswamy, J. & Gopal, S. (2013). Inbreeding effects on quantitative traits in random mating and selected populations of the mulberry silkworm, Bombyx mori. *J Insect Sci* **12**, 140.

Fan, H. W., Zhang, X. C., Xu, Y. P., Cheng, X. W. & Zhang, C. X. (2012). Genome of a Bombyx mori nucleopolyhedrovirus strain isolated from India. *J Virol* **86**, 11941.

Garavaglia, M. J., Miele, S. A., Iserte, J. A., Belaich, M. N. & Ghiringhelli, P. D. (2012). The ac53, ac78, ac101, and ac103 genes are newly discovered core genes in the family Baculoviridae. *J Virol* **86**, 12069-12079.

Gomi, S., Majima, K. & Maeda, S. (1999). Sequence analysis of the genome of Bombyx mori nucleopolyhedrovirus. *J Gen Virol* **80 (Pt 5)**, 1323-1337.

Gong, Y., Li, Z., Wang, L., Pan, L., Yang, K. & Pang, Y. (2003). Characterization of bro-b gene of Spodoptera litura multicapsid nucleopolyhedrovirus. *Virus Genes* **27**, 115-123.

Grace, T. D. (1967). Establishment of a line of cells from the silkworm Bombyx mori. *Nature* **216**, 613.

Guindon, S., Dufayard, J. F., Lefort, V., Anisimova, M., Hordijk, W. & Gascuel, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol* **59**, 307-321.

Harrison, R. L. & Popham, H. J. (2008). Genomic sequence analysis of a granulovirus isolated from the Old World bollworm, Helicoverpa armigera. *Virus Genes* **36**, 565-581.

Jehle, J. A., Blissard, G. W., Bonning, B. C., Cory, J. S., Herniou, E. A., Rohrmann, G. F., Theilmann, D. A., Thiem, S. M. & Vlak, J. M. (2006). On the classification and nomenclature of baculoviruses: a proposal for revision. *Arch Virol* **151**, 1257-1266.

Kang, W., Suzuki, M., Zemskov, E., Okano, K. & Maeda, S. (1999). Characterization of baculovirus repeated open reading frames (bro) in Bombyx mori nucleopolyhedrovirus. *J Virol* **73**, 10339-10345.

Katoh, K., Misawa, K., Kuma, K. & Miyata, T. (2002). MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res* **30**, 3059-3066.

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Meintjes, P. & Drummond, A. (2012). Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28**, 1647-1649.

Lung, O. Y., Cruz-Alvarez, M. & Blissard, G. W. (2003). Ac23, an envelope fusion protein homolog in the baculovirus Autographa californica multicapsid nucleopolyhedrovirus, is a viral pathogenicity factor. *J Virol* **77**, 328-339.

O'Reilly, D., Miller, L. K. & Luckow, V. A. (1992). *Baculovirus Expression Vectors: a laboratory manual*. New York: Freeman and Company.

Ogembo, J. G., Caoili, B. L., Shikata, M., Chaeychomsri, S., Kobayashi, M. & Ikeda, M. (2009). Comparative genomic sequence analysis of novel Helicoverpa armigera nucleopolyhedrovirus (NPV) isolated from Kenya and three other previously sequenced Helicoverpa spp. NPVs. *Virus Genes* **39**, 261-272.

Ono, C., Kamagata, T., Taka, H., Sahara, K., Asano, S. & Bando, H. (2012). Phenotypic grouping of 141 BmNPVs lacking viral gene sequences. *Virus Res* **165**, 197-206.

Pandiarajan, J., Cathrin, B. P., Pratheep, T. & Krishnan, M. (2011). Defense role of the cocoon in the silk worm Bombyx mori L. *Rapid Commun Mass Spectrom* **25**, 3203-3206.

Pang, M., Pan, G. Q., Li, T., Wang, X. & Zhou, Z. Y. (2007). [The analysis of bro genes of GD isolate of Bombyx mori nucleopolyhedrovirus]. *Bing Du Xue Bao* **23**, 485-489.

Peng, K., van Lent, J. W., Boeren, S., Fang, M., Theilmann, D. A., Erlandson, M. A., Vlak, J. M. & van Oers, M. M. (2012). Characterization of novel components of the baculovirus per os infectivity factor complex. *J Virol* **86**, 4981-4988.

Pereira, E. P., Conte, H., Ribeiro Lde, F., Zanatta, D. B., Bravo, J. P., Fernandez, M. A. & Brancalhao, R. M. (2008). Cytopathological process by multiple nucleopolyhedrovirus in the testis of Bombyx mori L., 1758 (Lepidoptera: Bombycidae). *J Invertebr Pathol* **99**, 1-7.

Pereira, N. C., Munhoz, R. E., Bignotto, T. S., Bespalhuk, R., Garay, L. B., Saez, C. R., Fassina, V. A., Nembri, A. & Fernandez, M. A. (2013). Biological and molecular characterization of silkworm strains from the Brazilian germplasm bank of Bombyx mori. *Genet Mol Res* **12**, 2138-2147.

Perera, O., Green, T. B., Stevens, S. M., Jr., White, S. & Becnel, J. J. (2007). Proteins associated with Culex nigripalpus nucleopolyhedrovirus occluded virions. *J Virol* **81**, 4585-4590.

Ribeiro Lde, F., Brancalhao, R. M., Torquato, E. F. & Fernandez, M. A. (2009). Susceptibility of the Bombyx mori cardia cells to Nucleopolyhedrovirus, multiple subgroup, BmMNPV. *J Invertebr Pathol* **100**, 195-198.

Rohrmann, G. F. (2013). *Baculovirus Molecular Biology*. Bethesda (MD): National Center for Biotechnology Information (US). Available from: http://www.ncbi.nlm.nih.gov/books/NBK49500/.

Rowley, D. L., Popham, H. J. & Harrison, R. L. (2011). Genetic variation and virulence of nucleopolyhedroviruses isolated worldwide from the heliothine pests Helicoverpa armigera, Helicoverpa zea, and Heliothis virescens. *J Invertebr Pathol* **107**, 112-126.

Simon, O., Palma, L., Beperet, I., Munoz, D., Lopez-Ferber, M., Caballero, P. & Williams, T. (2011). Sequence comparison between three geographically distinct Spodoptera frugiperda multiple nucleopolyhedrovirus isolates: Detecting positively selected genes. *J Invertebr Pathol* **107**, 33-42.

Slack, J. & Arif, B. M. (2007). The baculoviruses occlusion-derived virus: virion structure and function. *Adv Virus Res* **69**, 99-165.

Stamatakis, A., Hoover, P. & Rougemont, J. (2008). A rapid bootstrap algorithm for the RAxML Web servers. *Syst Biol* **57**, 758-771.

Summers, M. D. (2006). Milestones leading to the genetic engineering of baculoviruses as expression vector systems and viral pesticides. *Adv Virus Res* **68**, 3-73.

Vasyl'ieva, V. L. & Lebedynets, N. M. (2001). [Baculoviruses--a promising alternative to toxic chemicals in the battle against plant pests]. *Mikrobiol Z* **62**, 47-55.

Wang, R., Deng, F., Hou, D., Zhao, Y., Guo, L., Wang, H. & Hu, Z. (2010). Proteomics of the Autographa californica nucleopolyhedrovirus budded virions. *J Virol* **84**, 7233-7242.

Washburn, J. O., Chan, E. Y., Volkman, L. E., Aumiller, J. J. & Jarvis, D. L. (2002). Early synthesis of budded virus envelope fusion protein GP64 enhances Autographa californica multicapsid nucleopolyhedrovirus virulence in orally infected Heliothis virescens. *J Virol* **77**, 280-290.

Wf, I. J., van Strien, E. A., Heldens, J. G., Broer, R., Zuidema, D., Goldbach, R. W. & Vlak, J. M. (1999). Sequence and organization of the Spodoptera exigua multicapsid nucleopolyhedrovirus genome. *J Gen Virol* **80 (Pt 12)**, 3289-3304.

Xiang, X., Shen, Y., Yang, R., Chen, L., Hu, X. & Wu, X. (2013). Bombyx mori nucleopolyhedrovirus BmP95 plays an essential role in budded virus production and nucleocapsid assembly. *J Gen Virol* **94**, 1669-1679.

Xu, Y. P., Cheng, R. L., Xi, Y. & Zhang, C. X. (2013). Genomic diversity of Bombyx mori nucleopolyhedrovirus strains. *Genomics*.

Xu, Y. P., Ye, Z. P., Niu, C. Y., Bao, Y. Y., Wang, W. B., Shen, W. D. & Zhang, C. X. (2010). Comparative analysis of the genomes of Bombyx mandarina and Bombyx mori nucleopolyhedroviruses. *J Microbiol* **48**, 102-110.

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

Zhang, H., Yang, Q., Qin, Q. L., Zhu, W., Zhang, Z. F., Li, Y. N., Zhang, N. & Zhang, J. H. (2013). Genomic sequence analysis of Helicoverpa armigera nucleopolyhedrovirus isolated from Australia. *Arch Virol*.

9. Supplementary Materials

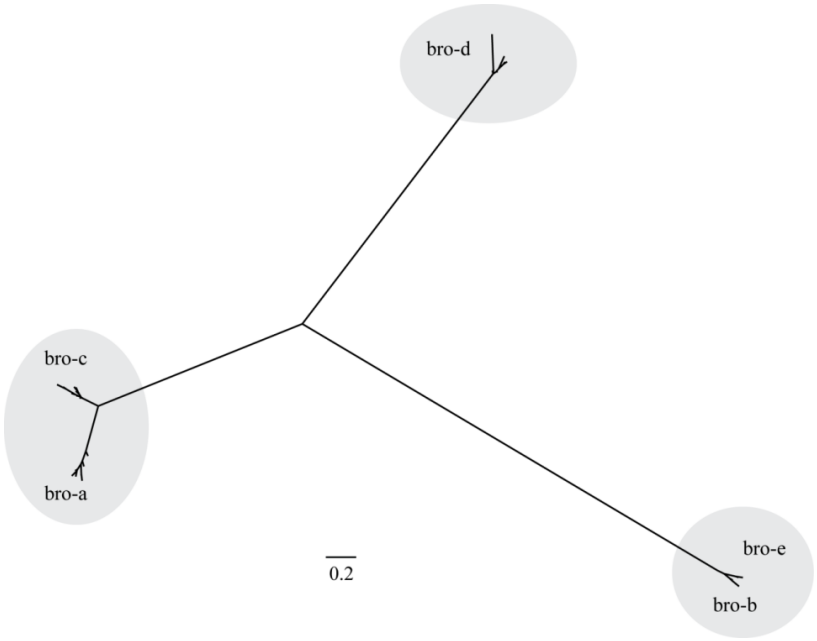

Fig. S1 Phylogenetic analysis of Bro proteins found in *Bombyx*-related viruses and AcMNPV. The maximum likelihood phylogenetic tree was inferred using a MAFFT alignment (25) of all Bro proteins and PhyML (27). The proteins clusters are highlighted in gray.

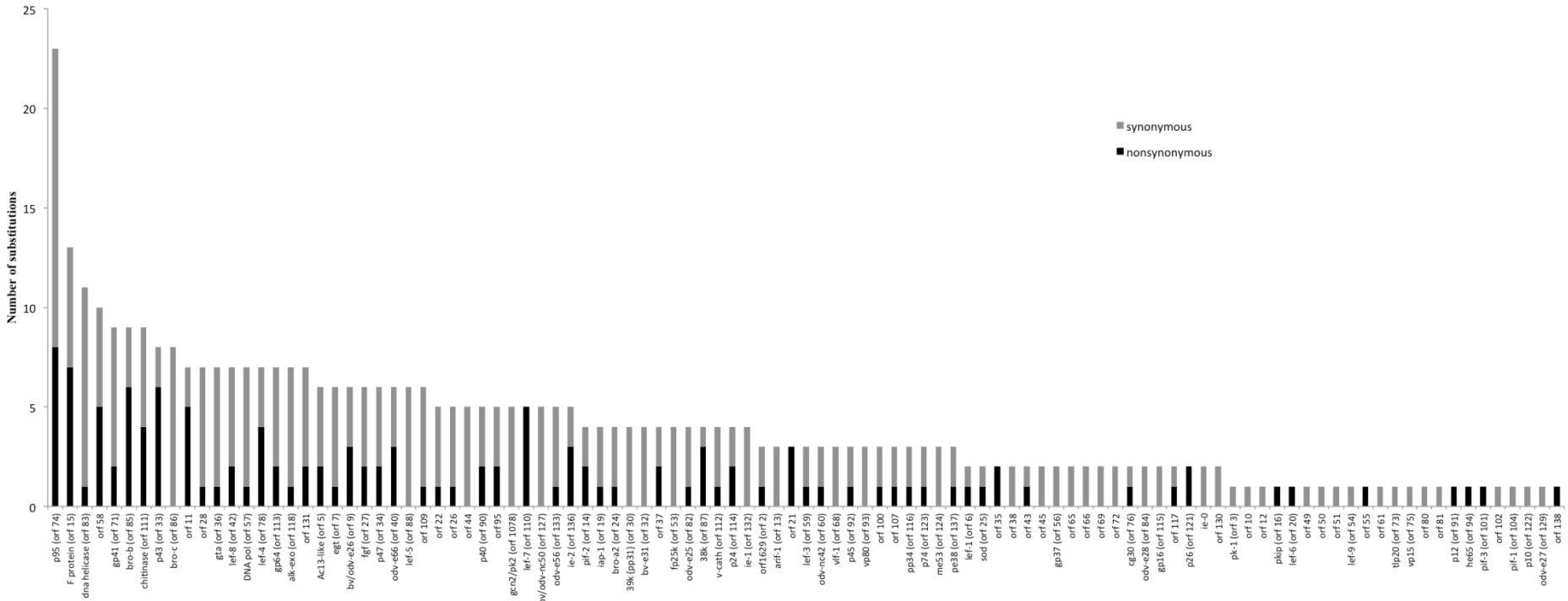

Fig. S2 Polymorphism distribution along the different genes of the Brazilian isolate. The number of polymorphisms is shown in the Y-axis. The synonymous changes are shown in gray and non-synonymous in black. The different genes are shown in the X-axis in decreasing order of polymorphisms. We included both common gene name (when present) and the orf number in the BmNPV-Brazilian genome.

Table S1. Characteristics of the BmNPV-Brazilian isolate genome. Predicted ORFs are compared with homologues in BmNPV-T3 and AcMNPV-C6.

Orf	Gene name	Position[a]	Size (aa)	BmNPV-T3		AcMNPV	
				ORF	Id[b] (%)	ORF	Id[b] (%)
1	*polh*	1 > 738	245	1	100	8	86.1
2	*orf1629*	768 < 2,387	539	2	95.9	9	86.5
3	*pk-1*	2,386 > 3,210	274	3	99.2	10	94.8
4		3,236 < 4,258	340	4	97.9	11	93.2
5		4,593 < 5,588	331	5	98.8	13	92.1
6	*lef-1*	5,468 < 6,280	270	6	98.9	14	94.1
7	*egt*	6,397 > 7,917	506	7	99.2	15	96
8		7,930 > 8,091	53	7a	100	-	-
9	*bv/odv-e26*	8,057 > 8,746	229	8	99.1	16	96.1
10		8,715 > 9,347	210	9	99.1	17	95.7
11		9,377 < 10,447	356	10	98.4	18	94.2
12		10,449 > 10,781	110	11	100	19	90.9
13	*arif-1*	10,968 < 12,284	438	12	96.4	20/21	88.5
14	*pif-2*	12,321 > 13,469	382	13	98.7	22	92.9
15	*f protein*	13,572 > 15,596	674	14	98.7	23	88.4
16	*pkip*	15,627 < 16,136	169	15	99.4	24	91.7
17	*dbp*	16,176 < 17,129	317	16	100	25	95.9
18		17,205 > 17,594	129	17	99.2	26	93.8
19	*iap-1*	17,596 > 18,471	291	18	94.9	27	92.4
20	*lef-6*	18,476 > 18,997	173	19	99.4	28	93.6
21		19,115 < 19,330	71	20	100	29	93
22		19,385 < 20,803	472	21	99.6	30	95.5
23	*bro-a1*	20,839 < 21,009	56	22	100	-	-
24	*bro-a2*	21,049 < 21,786	245		85.8	-	-
25	*sod*	21,908 > 22,363	151	23	97.4	31	96.7
26		22,732 > 22,974	80	-	-	-	-
27	*fgf*	22,931 > 23,479	182	24	97.3	32	90.2
28		24,004 < 24,651	215	25	96.7	34	94.4
29	*v-ubq*	24,672 > 24,905	77	26	100	35	100
30	*39k; pp31*	24,955 < 25,788	277	27	98.6	36	91
31	*lef-11*	25,782 < 26,120	112	28	98.2	37	97.3
32	*nudix*	26,083 < 26,736	217	29	100	38	96.3
33	*p43*	26,804 < 27,892	362	30	99.7	39	91.9
34	*p47*	27,900 < 29,099	399	31	99.7	40	97
35	*lef-12*	29,104 > 29,637	177	32	96	41	95.5
36	*gta*	29,713 > 31,233	506	33	100	42	96.4
37		31,247 > 31,483	78	34	98.8	43	91.1
38		31,464 > 31,859	131	35	100	44	98.5
39		31,861 > 32,445	194	36	99	45	89.2
40	*odv-e66*	32,430 > 34,556	708	37	98.4	46	93.3
41	*trax-like (ets)*	34,654 < 34,923	89	38	98.9	47	93.1
42	*lef-8*	35,168 < 37,798	876	39	99.8	48	97.7

43	*dna-J*	37,825	>	38,784	319	40	99.7	51	94.4
44		38,775	<	39,359	194	41	100	52	91.6
45		39,361	>	39,780	139	42	100	53	95.7
46	*lef-10*	39,777	>	40,013	78	42a	100	53a	96.2
47	*vp1054*	39,871	>	40,968	365	43	100	54	95.6
48		41,050	>	41,283	77	44	100	55	88.3
49		41,285	>	41,539	84	45	100	56	94
50		41,778	>	42,263	161	46	100	57	93.8
51	*chaB-like*	42,279	<	42,794	171	47	99.4	58/59	94.9
52	*chaB-like*	42,806	<	43,054	82	48	98.3	60	85.1
53	*fp-25k*	43,206	<	43,850	214	49	99.1	61	97.7
54	*lef-9*	43,954	>	45,426	490	50	99.6	62	98.8
55		45,487	>	45,954	155	51	98.7	63	92.3
56	*gp37*	46,029	<	46,913	294	52	99.3	64	95.6
57	*dna-pol*	47,203	<	50,022	939	53	99.6	65	96.5
58		50,031	>	52,448	805	54	98.6	66	90.2
59	*lef-3*	52,451	<	53,608	385	55	99.5	67	91.7
60		53,627	>	54,031	134	56	99.3	68	92.6
61	*mtase*	54,009	>	54,797	262	57	100	69	97.7
62	*iap-2*	54,946	>	55,695	249	58	100	71	95.6
63		55,754	>	55,936	60	58a	98.3	72	88.3
64		55,946	<	56,245	99	59	94.9	73	87.9
65		56,242	<	57,048	268	60	100	74	91.8
66		57,066	<	57,467	133	61	100	75	96.2
67		57,486	<	57,743	85	62	100	76	96.5
68	*vlf-1*	57,759	<	58,898	379	63	100	77	98.2
69		58,904	<	59,236	110	64	99.1	78	94.5
70		59,239	<	59,553	104	65	100	79	99
71	*gp41*	59,556	<	60,767	403	66	98.8	80	93.9
72		60,757	<	61,461	234	67	99.6	81	92.2
73		61,307	<	61,852	181	68	98.9	82	85
74	*p95 (vp91)*	61,818	>	64,331	837	69	98	83	90.7
75	*p15*	65,570	>	65,950	126	70	100	87	94.4
76	*cg30*	65,955	<	66,752	265	71	97.2	88	91.3
77	*vp39*	66,755	<	67,804	349	72	97.7	89	93.7
78	*lef-4*	67,823	>	69,220	465	73	99.4	90	96.6
79		69,217	<	69,681	154	74	100	91	51.1
80	*p33 (sox)*	69,718	<	70,497	259	75	100	92	97.3
81		70,496	>	70,981	161	76	100	93	98.8
82	*odv-e25*	70,990	>	71,676	228	77	99.6	94	90.8
83	*dna-helicase*	71,714	<	75,382	1222	78	99.8	95	95.9
84	*pif-4*	75,369	>	75,917	182	79	100	96	94.7
85	*bro-b*	76,013	>	76,738	241	80	93.4	-	-
86	*bro-c*	76,798	>	77,760	320	81	90.7	-	-
87	*38K*	77,907	<	78,869	320	82	100	98	93.4
88	*lef-5*	78,804	>	79,601	265	83	99.6	99	97.4

89	*p6.9*	79,598	<	79,795	65	84	98.5	100	78.5
90	*p40*	79,837	<	80,928	363	85	97.8	101	95.9
91	*p12*	80,948	<	81,325	125	86	97.6	102	95.6
92	*p45*	81,306	<	82,469	387	87	99.5	103	95.1
93	*vp80*	82,495	>	84,573	692	88	99.7	104	96.2
94	*he65*	84,595	<	85,464	289	89	100	105	95.1
95		86,173	>	86,922	249	90	100	106/107	87.9
96		86,899	<	87,240	113	91	100	108	96.2
97		87,255	<	88,430	391	92	100	109	96.2
98		88,454	<	88,633	59	92a	98.3	110	92.9
99		88,682	<	88,885	67	93	100	111	88.1
100		89,561	<	90,835	424	94	99.5	114	95.5
101	*pif-3*	90,857	<	91,471	204	95	99	115	92.7
102		91,465	<	91,638	57	95a	92.2	116	84.4
103		91,574	>	91,861	95	96	100	117	94.7
104	*pif-1*	91,991	>	93,574	527	97	99.4	119	84.6
105		93,582	>	93,830	82	98	98.8	120	92.7
106		93,933	>	94,106	57	98a	94.9	121	93
107		93,999	<	94,184	61	99	100	122	90.3
108	*pk-2*	94,218	<	94,895	225	100	99.1	123	95.8
109		95,079	>	95,813	244	101	99.2	124	87.1
110	*lef-7*	95,832	<	96,512	226	102	96.1	125	86.7
111	*chitinase*	96,502	<	98,160	552	103	99.6	126	94.7
112	*v-cath*	98,208	>	99,179	323	104	99.7	127	96.6
113	*gp64*	99,296	<	100,888	530	105	99.8	128	94.9
114	*p24*	101,015	>	101,602	195	106	99	129	90.9
115	*gp16*	101,630	>	101,950	106	107	100	130	100
116	*pp34*	102,012	>	102,953	313	108	98.7	131	88.3
117		102,956	>	103,618	220	109	99.6	132	95.5
118	*alk-exo*	103,646	>	104,908	420	110	99.8	133	95.5
119		105,023	>	105,235	70	111	100	-	-
120	*p35*	105,369	>	106,268	299	112	99.3	135	90.6
121	*p26*	106,996	>	107,718	240	113	98.8	136	93.3
122	*p10*	107,791	>	108,003	70	114	100	137	88.6
123	*p74*	108,089	<	110,026	645	115	99.5	138	90.9
124	*me53*	110,256	<	111,617	453	116	97.8	139	91
125		111,763	<	111,966	67	67[c]	100[c]	-	-
126	*ie-0*	111,894	>	112,679	261	117	100	141	96.9
127	*bv/odv-nc50*	112,694	>	114,124	476	118	100	142	98.5
128	*odv-e18*	114,132	>	114,440	102	119	88.2	143	83.9
129	*odv-e27*	114,455	>	115,327	290	120	100	144	99
130		115,342	>	115,629	95	121	100	145	93.5
131		115,624	<	116,229	201	122	98.6	146	96.6
132	*ie-1*	116,295	>	118,049	584	123	99.8	147	95.7
133	*odv-e56*	118,138	<	119,265	375	124	98.7	148	82.2
134		119,294	<	119,614	106	125	98.1	149	87.5

135		119,583	>	119,930	115	126	100	150	69.7
136	*ie-2*	119,963	<	121,243	426	127	97.7	151	71.5
137	*pe38*	121,731	>	122,660	309	128	98.4	153	83.3
138		122,762	>	122,995	77	129	98.8	154	81.8
139	*ptp-1*	123,682	>	124,188	168	130	100	1	97
140	*bro-d*	124,185	<	125,234	349	131	96.6	2	82
141		125,308	<	125,763	151	133	99.3	4	94
142		125,792	>	126,121	109	134	99.1	5	91.7
	lef-2	126,102	>	126,734	210	135	99.5	6	95.2

[a] Direction of putative transcripts is noted by the symbols > (sense) and < (antisense).
[b] Identity acquired from Psi-BLAST analysis.
[c] Even present in the BmNPV-T3, this ORF was not annotated.

Table S2. Characteristics of the homolog regions (HRs) in *Bombyx*-isolated baculoviruses.

Name[1]	Id (%)[2]	Size ± SD[3] (bp)	HR Size (bp)										
			BmNPV									BomaNPV	
			Brazilian	T3	Guangxi	Zheijang	India	C1	C2	C6	Cubic	S1	S2
HR1	89.5	540.5 ± 46.6	527	592	582	527	594	566	458	458	527	584	527
HR2L	86.8	622.0 ± 183.2	251	604	-	-	620	918	513	513	784	611	784
HR2R	90.4	255.6 ± 28	258	267	267	258	168	258	267	267	267	268	267
HR3	94.2	498.8 ± 80.1	549	549	547	534	548	553	553	345	381	547	381
HR4L	74.9	361.6 ± 75.5	437	218	361	434	505	289	361	289	361	362	361
HR4R	95.7	582.8 ± 25.9	592	591	501	591	591	590	591	591	591	591	591
HR5	88.3	601.2 ± 68.9	550	615	552	552	726	659	659	659	553	615	473

1. Based on BmNPV-T3 nomenclature (12). 2. Global pairwise identity obtained by MAFFT alignment. 3. Standard deviation.

Capítulo 3. Genome sequence of *Erinnyis ello granulovirus* (ErelGV), a natural cassava hornworm pesticide and the first sequenced sphingid-infecting betabaculovirus

1. Abstract

Background. Cassava (*Manihot esculenta*) is the basic source for dietary energy of 500 million people in the world. In Brazil, *Erinnyis ello ello* (Lepidoptera: Sphingidae) is a major pest of cassava crops and a bottleneck for its production. In the 1980s, a naturally occurring baculovirus was isolated from *E. ello* larva and successfully applied as a bio-pesticide in the field. Here, we described the structure, the complete genome sequence, and the phylogenetic relationships of the first sphingid-infecting betabaculovirus. **Results.** The baculovirus isolated from the cassava hornworm cadavers is a betabaculovirus designated *Erinnyis ello granulovirus* (ErelGV). The 102,759 bp long genome has a G+C content of 38.7%. We found 130 putative ORFs coding for polypeptides of at least 50 amino acid residues. Only eight genes were found to be unique. ErelGV is closely related to ChocGV and PiraGV isolates. We did not find typical homologous regions and *cathepsin* and *chitinase* homologous genes are lacked. The presence of *he65* and *p43*homologous genes suggests horizontal gene transfer from *Alphabaculovirus*. Moreover, we found a nucleotide metabolism-related gene and two genes that could be acquired probably from *Densovirus*. **Conclusions.** The ErelGV represents a new virus species from the genus *Betabaculovirus* and is the closest relative of ChocGV. It contains a *dUTPase*-like, a *he65*-like, *p43*-like genes, which are also found in several other alpha- and betabaculovirus genomes, and two *Densovirus*-related

genes. Importantly, recombination event between insect viruses from unrelated families and genera might drive baculovirus genomic evolution.

Key-words: biological control, cassava hornworm, baculovirus, Sphingidae, horizontal gene transfer, *Betabaculovirus* evolution.

Este copítulo foi inteiramente publicado na revista BMC genomics. **Ardisson-Araujo, D. M., de Melo, F. L., Andrade Mde, S., Sihler, W., Bao, S. N., Ribeiro, B. M. & de Souza, M. L. (2014).** Genome sequence of Erinnyis ello granulovirus (ErelGV), a natural cassava hornworm pesticide and the first sequenced sphingid-infecting betabaculovirus. *BMC Genomics* **15**, 856.

2. Background

Cassava (*Manihot esculenta*) is the basic source for dietary energy of 500 million people in tropical and subtropical areas of Africa, Asia, and Latin America (El-Sharkawy, 2004). In Brazil, the hornworm *Erinnyis ello ello*(Lepidoptera: Sphingidae) is one of the most important pests (Pietrowski *et al.*, 2010) occurring throughout the year and impacting greatly cassava production (Bellotti *et al.*, 1992; Fazolin *et al.*, 2007). This pest has been observed in 35 plant species, especially in the Euphorbiaceae family. In large infestations, the cassava pest may reduce by 50% the roots yield. In the 1980s, a naturally occurring baculovirus was isolated from this pest and applied as a bio-pesticide in Brazil (Schmitt, 1985). The biological control program has proven to be safe and economical (Schmitt, 1985; Schmitt, 2002). However, genomic and structural information about the virus is lacking.

The *Baculoviridae* is a family of insect viruses with circular double-stranded genomic DNA (Herniou *et al.*, 2012; Jehle *et al.*, 2006a; Rohrmann, 2013) that have been successfully applied in controlling agricultural and forest pests (Moscardi, 1999). So far, *Alpha* and *Betabaculovirus* are the most studied baculovirus genera; both infect Lepidoptera (Rohrmann, 2013). The infection is initiated when larvae feed on foliage contaminated with orally infectious occlusion bodies (OBs) (Ji *et al.*, 2010) that release occlusion derived-virions (ODVs) in the midgut (Slack & Arif, 2007). Early after primary midgut epithelial cell infection, budded virions (BV) are produced and cause systemic infection. Infection symptoms include cuticle discoloration, movement loss, and incapability for feeding (Wang *et al.*, 2010b; Washburn *et al.*, 2003).

Few full-length betabaculovius genome sequences are available compared to those from *Alphabaculovirus* and none of them was isolated from sphingid host. In this context, identification and sequencing of virus species from different lepidopteran families will provide a wider empirical database to help understand baculovirus evolution (Herniou *et al.*, 2001; Herniou *et al.*, 2003). Here, we presented the morphological characterization, the complete genome sequence, and the phylogenetic analyses of the natural cassava hornworm pesticide, the first completely sequenced betabaculovirus isolated from a sphingid host.

3. Results and Discussion

3.1. Virus characterization and genome features

A naturally occurring baculovirus was isolated from dead cassava hornworm (*E. ello ello*) caterpillars in crops from the South of Brazil in 1986. As shown in Figure 1A, the larvae is usually found hanged in cassava apical leaves, which is a characteristic symptom of the baculovirus infections(Hoover *et al.*, 2011). Neither cuticle melanization nor post-mortem melting phenotypes were observed among the caterpillar cadavers, an attribute which probably facilitated virus collection and use for pesticide production as previously observed in another baculovirus (*Anticarsia gemmatalis multiple nucleopolyhedrovirus* - AgMNPV) (Moscardi, 1999). Ultrastructural analyses revealed a granular OB with irregular form and size (Figure 1B) containing single rod-shaped nucleocapsid (Figure 1C). Both of these structural features, *i.e.* granular form and nucleocapsid shape, are typical of viruses from the genus *Betabaculovirus* (Ackermann & Smirnoff, 1983; Jehle *et al.*, 2006a) and thus, we named it *Erinnyis ello granulovirus* (ErelGV) isolate Br-S86 (Brazil/South/1986). Two other cassava hornworm-isolated granuloviruses were previously reported, one isolated in Colombia (Finnerty *et al.*, 2000) and another from an undisclosed geographical source (Jehle *et al.*, 2006b). Restriction endonuclease profile analyses (Figure 1D) suggest that the Brazilian and the Colombian viruses (previously published in (Finnerty *et al.*, 2000)) are either variants of the same species or are distinct species infecting the same host. However, the absence of sequence data from the latter prevents establishment of any phylogenetic relationship.

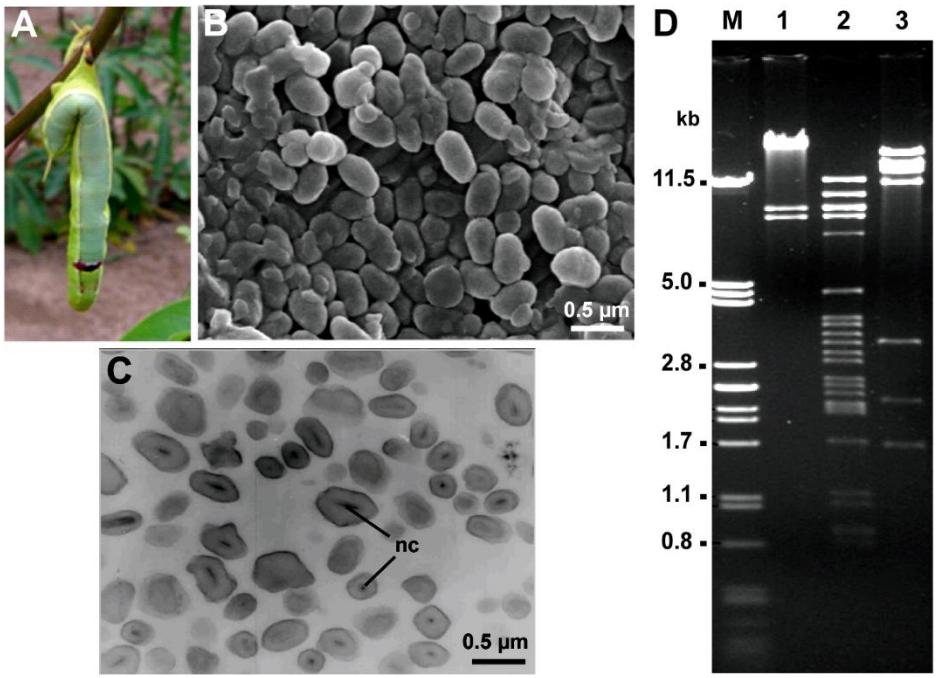

Figure 1 - *Erinnyis ello granulovirus* (ErelGV) infection and virus characterization. (A) Cassava hornworm cadaver found hanging in the field due to terminal baculovirus infection (Source: José Osmar Lorenzi). (B) Scanning and (C) transmission electron micrographs reveal granular occlusion bodies containing singly embedded rod-shaped nucleocapsid (nc) (scale bars = 0.5 µm). (C) Restriction enzyme profile of Brazilian isolate genomic DNA. Agarose gel electrophoresis-resolved DNA fragments digested with *Hin*dIII (lane 1), *Eco*RI (lane 2), *Bam*HI (lane 3).

We sequenced the genome of ErelGV, the first completely sequenced sphingid host-isolated betabaculovirus (Genbank accession number KJ406702). The genome is 102,759 bplong with a G+C content of 38.7% (Table 1). We found 130 putative genes coding for polypeptides of at least 50 amino acid residues. Table S1 summarizes the ErelGV genes and compares each predicted protein sequence with its orthlogues in other baculoviruses. Eight of these were shown to be unique (*ErelOrf-11*, *ErelOrf-15*, *ErelOrf-27*, *ErelOrf-53*, *ErelOrf-59*, *ErelOrf-70*, *ErelOrf-90*, *ErelOrf-102*) (Figure 3, in

red), and all of them are peptides with no significant similarity to any other sequence in GenBank. All 37 *Baculoviridae* core genes were found (Figure 3, in boldface). We identified five putative homologous regions (hrs)/repeat regions lacking typical alphabaculovirushr palindromes. This feature is also found in both *Choristoneura occidentalis granulovirus* (ChocGV) and *Pieris rapae granulovirus* (PiraGV) genomes. As observed in ChocGV (Escasa *et al.*, 2006), ErelGV lacks both *gp37* and *exon0*, which was previously predicted for being shared among all *Alpha* and *Betabaculovirus* (Garavaglia *et al.*, 2012).

Table 1. All species from the genus*Betabaculovirus* completely sequenced to date.

Virus species	Host Family	Size (bp)	ORFs	Accession	Refs.
Adoxophyesorana granulovirus	Tortricidae	99,657	119	AF547984	(Wormleaton *et al.*, 2003)
*Agrotissegetum granulovirus*Xinjiang	Noctuidae	131,680	132	AY522332	(Zhang *et al*., 2014)
*Agrotissegetum granulovirus*L1	Noctuidae	131,442	149	KC994902	(Zhang *et al*., 2014)
Choristoneuraoccidentalis granulovirus	Tortricidae	104,710	116	DQ333351	(Escasa *et al*., 2006)
Closteraanachoreta granulovirus	Notodontidae	101,487	123	HQ116624	(Liang *et al*., 2011)
Clostera anastomosis L. granulovirus	Notodontidae	101,818	123	KC179784	u/d
Cryptophlebialeucotreta granulovirus	Tortricidae	110,907	129	AY229987	(Lange & Jehle, 2003)
Cydiapomonella granulovirus	Tortricidae	123,500	143	U53466	(Luque *et al*., 2001)
Epinotiaaporema granulovirus	Tortricidae	119,092	132	JN408834	(Ferrelli *et al.*, 2012)
Erinnyisello granulovirus	Sphingidae	102,759	135	KJ406702	-
Helicoverpaarmigera granulovirus	Noctuidae	169,794	179	EU255577	(Harrison & Popham, 2008)
Phthorimaeaoperculella granulovirus	Gelechiidae	119,217	130	AF499596	u/d
*Pierisrapae granulovirus*China	Pieridae	108,592	120	GQ884143	(Zhang *et al*., 2012)
Pierisrapae granulovirus E3	Pieridae	108,476	125	GU111736	u/d
Pierisrapae granulovirus South Korea	Pieridae	108,658	120	JX968491	u/d
Plutellaxylostella granulovirus	Plutellidae	100,999	120	AF270937	(Hashimoto *et al.*, 2000)
Pseudaletiaunipuncta granulovirus	Noctuidae	176,677	183	EU678671	u/d
Spodopteralitura granulovirus	Noctuidae	124,121	136	DQ288858	(Wang *et al.*, 2011)
Xestia c-nigrum granulovirus	Noctuidae	178,733	181	AF162221	(Hayakawa *et al.*, 1999)

u/d - unpublished data

3.2. Phylogenetic analysis

In order to better understand the evolutionary history of ErelGV and the genus *Betabaculovirus*, we carried out a maximum likelihood phylogenetic analysis using the

37 baculovirus core gene alignment from all baculovirus genome available. ErelGV clustered with ChocGV and both viruses share the same ancestor with PiraGV isolates (Figure 2). Since the Chinese and Korean PiraGV isolates (Table 1) are very similar to each other (99.5%), we have included only the Chinese isolate in our analyses. Using Mauve alignment (Darling *et al.*, 2004), we found that ChocGV and PiraGV genomes have respectively 38.5% and 34.5% of global pairwise identity when compared to ErelGV genome. Additionally, our phylogenetic analyses did not find support for*Betabaculovirus* division in two clades (A and B), as described previously using neighbor joining clustering method (Ferrelli *et al.*, 2012; Liang *et al.*, 2011). Phylogenetic relationships in *Baculoviridae*, in particular in the genus *Betabaculovirus*, are difficult to discern due to the limited number of sequenced genomes available (Table 1).Therefore, we further evaluated ErelGV phylogenetic relationships using *granulin*, *lef-8*, and *lef-9*partial gene dataset as previously carried out (Jehle *et al.*, 2006b; Lange *et al.*, 2004) (28 partial sequences), but including new sequences publicly available (seven sequences from completely sequenced baculovirus) totalizing 35 granulovirus sequences. This analysis revealed that ErelGV isolate Br-S86 is closely related to another ErelGV (also called EeGV) from the Steinhaus collection (Jehle *et al.*, 2006b) and that both are closer to *Andraca bipunctata granulovirus* (AnbiGV) (data not shown).

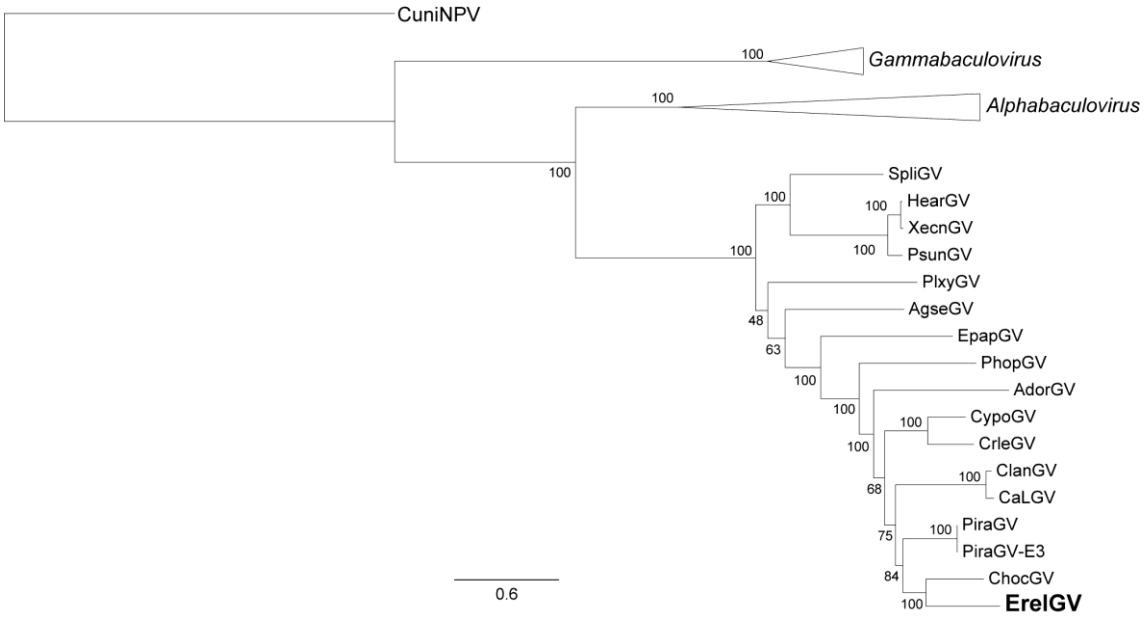

Figure 2 - Maximum likelihood tree for *Betabaculovirus*. The phylogenetic inference was based on the concatenated amino acid sequences of the 37 core genes identified in all complete baculovirus genome sequences. We collapsed all the *Gammabaculovirus* and *Alphaphabaculovirus*. The CuniNPV was used as root. ErelGV (boldface) clustered with ChocGV and both were closely related to PiraGV isolates.

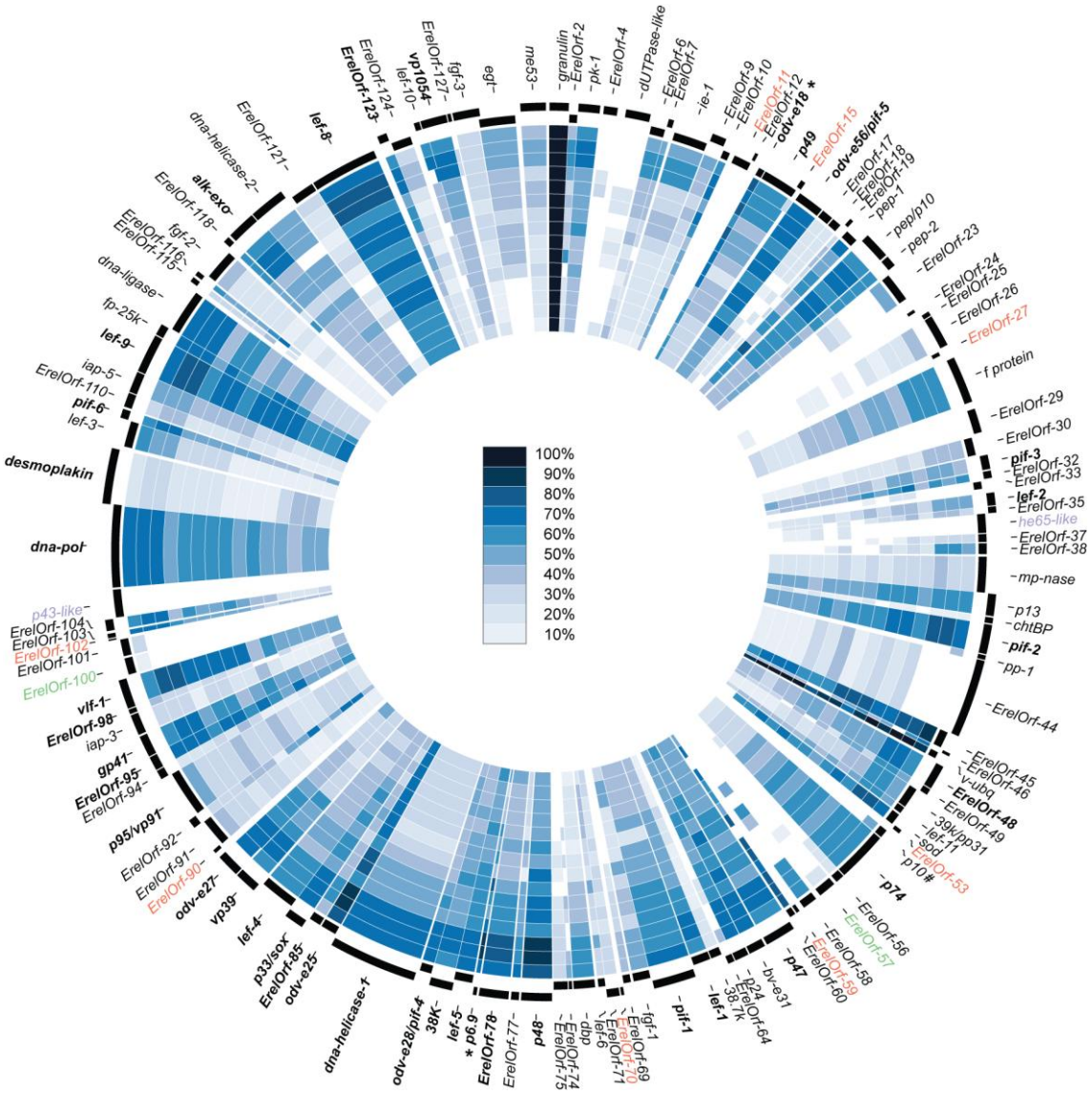

Figure 3 - Gene comparison of ErelGV genome and all completely sequenced betabaculoviruses available in Genbank. CDS identities were acquired by BLAST analysis and ranked from 0 to 100%. From the outermost ring: ChocGV, PiraGV-E3, PiraGV-China, ClanGV, CaLGV, CrleGV, CypoGV, AdorGV, PhopGV, EpapGV, AgseGV, PlxyGV, PsunGV, XecnGV, HearGV, and SpliGV-K1. For this representation, gene synteny is not taken into account. CDS that were absent in the ErelGV genome but present in the query sequences were not displayed. To prevent the missing of known homologues, like p6.9 and odv-e18 (asterisk), all the low identity hits

(bellow 20%) were plotted as well. Unique genes are shown in red, core genes are in boldface, and *Densovirus*-related genes are shown in green.

3.3. *Betabaculovirus* gene comparison

We performed BLAST comparisons between ErelGV and all other full betabaculovirus genomes available in Genbank using the CGView Comparison Tool (Grant *et al.*, 2012) and CIRCOS (Krzywinski *et al.*, 2009). As shown in Figure 3, most of the ErelGV-encoded ORFs are conserved among all betabaculovirus, but protein similarity varies widely across the species. Some structural proteins, such as granulin and the *per os* infectivity factors (PIFs), were the most conserved genes. Conversely, F protein, the major *Betabaculovirus* envelope fusion protein (EFP, encoded by *ErelOrf-28*) and matrix metalloproteinase (MMP, a stromelysin-1-like protein, encoded by *ErelOrf-39*) were particularly variable despite of both being present in every betabaculovirus sequenced to date. The EFP is essential for cell-to-cell movement and systemic virus spread (Rohrmann, 2013).GP64 is the EFP found in Group I *Alphabaculovirus* and all orthologues are closely related to each other (81 % of protein sequence identity), whereas the F protein, found in both *Alpha* and *Betabaculovirus* (Pearson & Rohrmann, 2002), is very diverse (20 to 40% sequence identity). Interestingly, deletion of the *gp64* or *f protein* genes is lethal for BV propagation in *Autographa californica multiple nucleopolyhedrovirus* (AcMNPV) (Oomens & Blissard, 1999) and *Helicoverpa armigera nuclepolyhedrovirus* (HaNPV) (Wang *et al.*, 2008; Wang *et al.*, 2010a), respectively. The deficiency can be rescued by *efp* homologous from many different viruses in the case of AcMNPV (Lung *et al.*, 2002), but the opposite is not true; AcMNPV *gp64* is not able to completely rescue an *f protein*-deleted HaNPV. However,

it is not clear why the F protein from *Plutella xylostella granulovirus* (PlxyGV) is not able to rescue the infectivity of *gp64*-null AcMNPV (Lung *et al.*, 2002) but that from AgseGV can. PlxyGV causes systemic infection to the diamondback moth *P. xylostella* (Plutellidae) larvae (Harrison & Lynn, 2007) and AgseGV infects the cutworm *A. segetum* (Noctuidae) (Wennmann & Jehle, 2014). Thus, the betabaculovirus EFP variability might reflect the cell machinery adjustment at the insect family level considering that AcMNPV infects caterpillar from the same insect family of *A. segetum*. A second highly variable gene, MMP, is a proteinase able to produce a distinct pattern of melanization in *Bombyx mori* larvae infected with the *Xestia c-nigrum granulovirus* (XecnGV) metalloproteinase-expressing *Bombyx mori nucleopolyhedrovirus* (Ko *et al.*, 2000). The enzyme is thought to enhance, replace, or act synergistically with proteins from virus or host playing an important role in the virus spread (Means & Passarelli, 2010). This variability is not unexpected since granulovirus genomes vary in content with respect to the presence or absence of the proteases *cathepsin* and *enhancin* genes and also the *chitinase* gene, which seemingly converge a redundant enzymatic activity but not necessarily function(Ko *et al.*, 2000; Lepore *et al.*, 1996; Means & Passarelli, 2010; Slack & Arif, 2007).

3.4. Lack of *cathepsin* and *chitinase* genes

ErelGV lacks *cathepsin* and *chitinase* genes, despite of their importance for promoting baculovirus horizontal transmission (D'Amico *et al.*, 2013). This feature can explain the integrity of caterpillar flesh and light color after death (Figure 1A). Other betabaculovirus genomes also lack both enzymes: complete deletion in ChocGV (Escasa *et al.*, 2006), *Adoxophyes orana granulovirus* (AdorGV) (Wormleaton *et al.*,

2003), *Phthorimaea operculella granulovirus* (PhopGV) (unpublished), PlxyGV(Hashimoto *et al.*, 2000) and *Spodoptera litura granulovirus* (SpliGV)(Wang *et al.*, 2011); *Cryptophlebia leucotreta granulovirus* (CrleGV)(Lange & Jehle, 2003) *chitinase* has an interruption; and in *Helicoverpa armigera granulovirus* (HearGV) (Harrison & Popham, 2008) only *cathepsin* is absent. Interestingly, most of these deletions seem to have occurred independently of each other within *Betabaculovirus* (data not shown), aside from ChocGV and ErelGV in which is strongly supported an ancestral lacking. Thus, it is reasonable to expect that AnbiGV, the closest relative to ErelGV, might also lack both *cathepsin* and *chitinase*. Taken together, these results reinforce the notion that both genes are most likely non-essential for the persistence of baculoviruses in the environment. Conversely, previous work from our research team has shown that introduction of *cathepsin* and *chitinase* from *Choristoneura fumiferana defective nucleopolyhedrovirus* into AgMNPV (which naturally lacks both genes) increases pathogenicity and occlusion body production relative to the wild type virus (Lima *et al.*, 2013).

3.5. *dUTPase-like* gene

ErelOrf-5 codes for a nucleotide metabolism-related gene homologous to *Orgyia pseudotsugata multiple nucleopolyhedrovirus* (OpMNPV) *Orf-31*. The gene seems to be composed of a fusion between two distinct ORFs; a baculovirus *thymidylate kinase*-like gene and *dUTPase*-like genes. The thymidylate kinase enzyme catalyzes a critical step in the biosynthesis of deoxythymidine triphosphate (Cui *et al.*, 2013). dUTPase catalyses dUTP dephosphorylation to generate dUMP (Penades *et al.*, 2013). High levels of dUTP can be deleterious for virus genomic DNA replication since dTTP can

be substituted for dUTP during DNA synthesis (Priet *et al.*, 2006). A high dUTP/dTTP ratio promotes uracil incorporation into DNA. Uracils in DNA are then targeted by uracil DNA glycosylase and excised, leading to futile repair cycles and DNA breakage and or translesional DNA synthesis (Castillo-Acosta *et al.*, 2012; Guillet *et al.*, 2006). Nucleotide metabolism-related enzyme acquisition is common in baculoviruses (Ferrelli *et al.*, 2012) and could avoid this deleterious response by decreasing the dUTP/dTTP ratio, however how these genes alter the virus fitness is not clear (Herniou & Jehle, 2007).

3.6. The *he65*-like and *p43*-likehomologues

The ErelGV genome contains homologues of the *he65* and *p43* genes. Homologues of *he65*are harbored by several alphabaculoviruses, four betabaculoviruses (*Agrotis segetum granulovirus* (AgseGV), HearGV, *Pseudaletia unipuncta granulovirus* (PsunGV), and *Xestia c-nigrum granulovirus* (XecnGV)), and two betaentomopoxviruses (*Amsacta moorei entomopoxvirus* - AMV and *Mythimna separate entomopoxvirus* - MSV). This gene is a member of a distinct RNA ligase family related to the *T4 RNA ligase gp63*-like gene and is present in all the domains of life (Bacteria, Archaea, and Eukarya) (Ho & Shuman, 2002; Rohrmann, 2013). The alignment of baculovirus and entomopoxvirus*he65*-like genes revealed large, independent, and recurrent deletions in the C-terminal region (data not shown), which contain five nucleotidyl transferase motifs (Ho & Shuman, 2002). The amino-terminal region was highly conserved although no previously characterized motifs were present. We performed a phylogenetic reconstruction based on this conserved domain. The *he65* reconstruction revealed distinct horizontal gene transfer (HGT) events from

Alphabaculovirus to *Betabaculovirus* and *Betaentomopoxvirus* (Figure 4A). *Betabaculovirus* likely endured two independent acquisitions from Group II *Alphabaculovirus* in distinct genomic regions: (*i*) a synapomorphic introduction for HearGV, PsunGV, and XecnGV (Figure 4A, yellow rectangle); and (*ii*) an additional gain for AgseGV (Figure 4A, brown rectangle). Importantly, support for AgseGV branch is low. However, the genomic context of the gene is conserved among HearGV, PsunGV, and XecnGV but not in AgseGV (data not shown), reinforcing our hypothesis that two independent introductions occurred. Likewise, *Betaentomopoxvirus* homologues were probably acquired from Group II *Alphabaculovirus* (Figure 4, orange rectangle). Remarkably, ErelGV is the first *Betabaculovirus* with a *he65*-like gene (*ErelOrf-36* - Figure 4A, purple rectangle) acquired from Group I *Alphabaculovirus*. It is not clear whether C-terminal deleted he65 remains functional in baculovirus. However the maintenance of the amino-terminal region indicates that this gene region is under positive selection pressure.

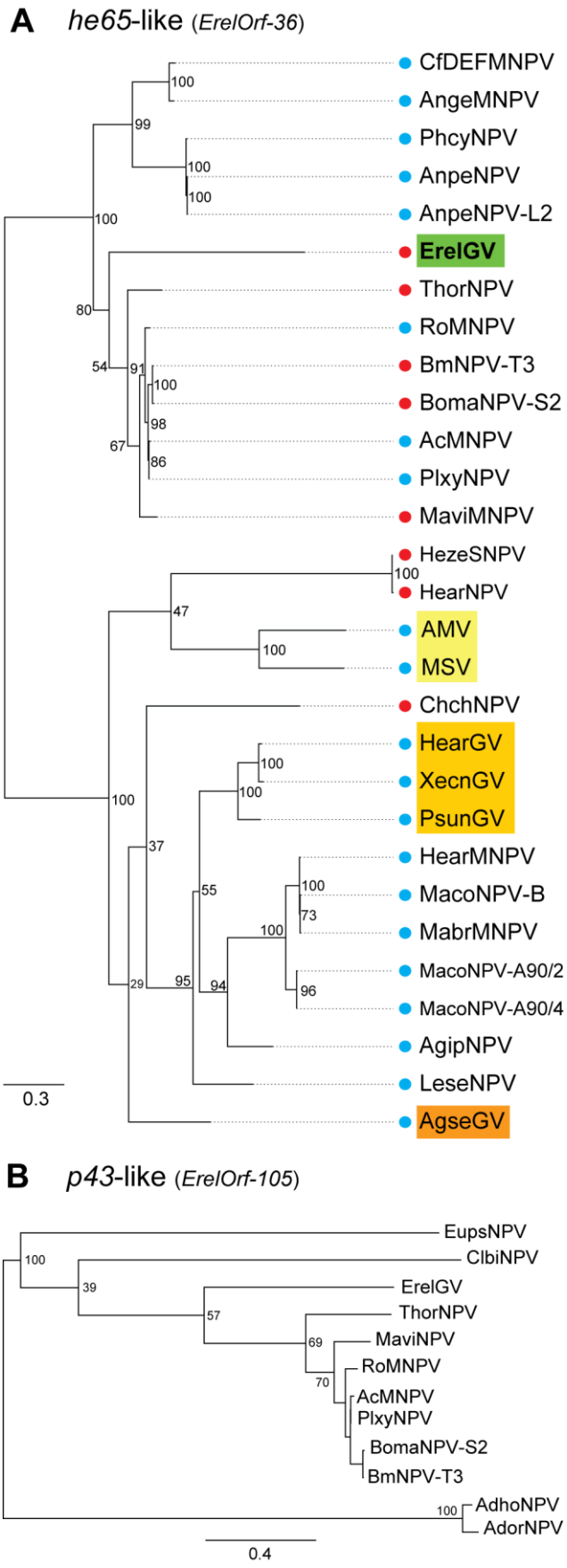
A
he65-like (ErelOrf-36)
CfDEFMNPV
AngeMNPV
PhcyNPV
AnpeNPV
AnpeNPV-L2
ErelGV
ThorNPV
RoMNPV
BmNPV-T3
BomaNPV-S2
AcMNPV
PlxyNPV
MaviMNPV
HezeSNPV
HearNPV
AMV
MSV
ChchNPV
HearGV
XecnGV
PsunGV
HearMNPV
MacoNPV-B
MabrMNPV
MacoNPV-A90/2
MacoNPV-A90/4
AgipNPV
LeseNPV
AgseGV
0.3
B
p43-like (ErelOrf-105)
EupsNPV
ClbiNPV
ErelGV
ThorNPV
MaviNPV
RoMNPV
AcMNPV
PlxyNPV
BomaNPV-S2
BmNPV-T3
AdhoNPV
AdorNPV
0.4

Figure 4 - Phylogeny of *he65*and *p43* reveals horizontal gene transfer in ErelGV from *Alphabaculovirus*. (A) The maximum likelihood (ML) tree was inferred using the conserved amino-terminal region alignment of *he65*-like gene for 36 baculoviruses and two entomopoxviruses. Circles indicate the presence (blue) or absence (red) of the carboxy-terminal region. The postulated horizontal gene transfer (HGT) events are highlighted for *Betabaculovirus* (light and dark orange), *Betaentomopoxvirus* (yellow), and ErelGV (green). (B) ML-Phylogenetic reconstruction for *p43*-like gene found in ErelGV genome. The trees are midpoint rooted for purposes of clarity.

Furthermore, we found in ErelGV genome a *p43*-like gene (*ErelOrf-105*) whose homologues were found only in baculovirus species from the genus *Alphabaculovirus* with conserved amino acid sequence and position in the genome (Rohrmann, 2013). Deletion of *p43* in AcMNPV does not affect virus replication in cell culture and the reason for gene acquisition and preservation is not clear (Yu & Carstens, 2011). Two hypotheses can be raised for *p43* introduction in ErelGV: (*i*) ErelGV acquired the *p43*-like gene from Group I *Alphabaculovirus*, specifically from AcMNPV-related viruses; or (*ii*) ErelGV acquired from Group II *Alphabaculovirus*, specifically from a baculovirus (*e.g. Clanis bilineata nucleopolyhedrovirus* - ClbiNPV (Zhu *et al.*, 2009)) during co-infection of a sphingid host.

3.7. Acquisitions of *Densovirus*-related genes in *Betabaculovirus*

ErelOrf-57 and *ErelOrf-100* are homologues to a non-structural *Densovirus* gene. *Densovirus*-related genes were previously described in two betabaculoviruses, ChocGV (*ChocOrf-25*) (Escasa *et al.*, 2006) and CrleGV (*CrleOrf-9*) (Lange & Jehle, 2003), and

one gammabaculovirus (baculovirus infective to hymenoptera), *Neodiprion lecontei nucleopolyhedrovirus* (NeleNPV - *NeleOrf-81*) (Lauzon *et al.*, 2004). The latter did not match the other two homologues (data not shown), suggesting these genes resulted from at least two HGT events between densoviruses and baculoviruses. Despite the limited number of *Densovirus* genomes available, we performed a phylogenetic analysis to help understand the origins of *Betabaculovirus* homologues. We found that the genes were dispersed over the phylogenetic tree, suggestive of multiple HGT events. As shown in Figure 5, *Betabaculovirus* homologues did not form a monophyletic cluster. To further substantiate our findings, we compared the likelihood of the observed tree to that estimated assuming a *Betabaculovirus* monophyletic clade (single-HGT event). Indeed, the likelihood ratio test rejected the monophyletic hypothesis favoring the multiple-HGT scenario, which was also supported by the distinct genomic context observed for the homologous betabaculovirus genes (data not shown). Moreover, both *ErelOrf-57* and *ErelOrf-100* form a well-supported clade, indicating that they probably represent a gene duplication event during ErelGV evolution.

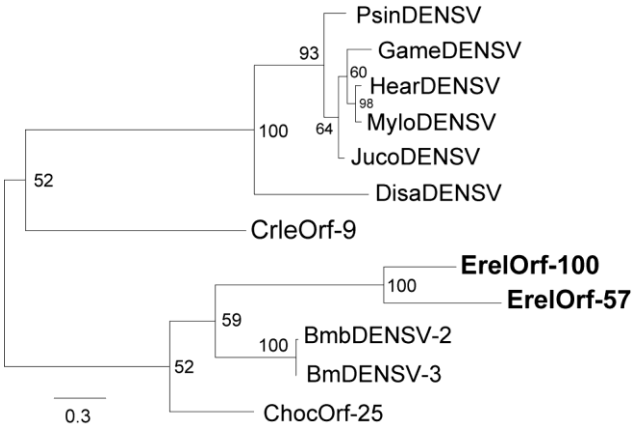

Figure 5. *Densovirus*-related genes in betabaculovirus and phylogenetic relationship. ML tree was inferred using the alignment of ErelOrf-57 and ErelOrf-100

from ErelGV with non-structural protein (NS) from *Bombyx mori densovirus 2* and *3* (BmbDENSV-2 / Genbank YP_007714627.1 and BmDENSV-3 / Genbank YP_007714627), NS3 from *Diatraea saccharalis densovirus* (DisaDENSV / Genbank NP_046812.1), *Mythimna loreyi densovirus* (MyloDENSV / Genbank NP_958098.1), *Helicoverpa armigera densovirus* (HaDENSV / Genbank AFK91982.1), *Galleria mellonella densovirus* (GameDENSV / Genbank NP_899649.1), *Junonia coenia densovirus* (JucoDENSV / Genbank AGO32182.1), and *Pseudoplusia includens densovirus* (PsinDENSV / GenbankYP_007003822.1), *Orf-25* from ChocGV, and *Orf-9* from CrleGV. The tree is midpoint rooted for purposes of clarity only. We hypothesized gene duplication for both ErelGV genes (boldface).

4. Conclusion

ErelGV is a new betabaculovirus species closely related to ChocGV and PiraGV isolates. Its genome encodes 130 ORFs, eight of which are unique. We found evidence suggesting horizontal gene transfers from *Alphabaculovirus* and *Densovirus* to *Betabaculovirus*. The *he65*-like gene was independently acquired three times from *Alphabaculovirus*. We found a *dUTPase* homologous and two *Densovirus*-related genes.The contribution of these genes to baculovirus fitness is not clear and is being experimentally tested in our lab**.** Importantly, recombination event between insect viruses from unrelated families and genera might drive baculovirus genomic evolution.

5. Material and Methods

5.1. Virus purification

Insect cadavers of the hornworm *E. ello ello* with baculovirus infection symptoms were collected in cassava crops in the South of Brazil (Itajaí, Santa Catarina) in 1986. They were kindly provided by Dr. Renato Arcanjo Pegoraro (EPAGRI). The cadavers were kept in the freezer and later used for OBs purification. Insect cadavers were homogenized with ddH_2O (w/v), filtered through three layers of gauze, and centrifuged at 7,000 *x g* for 10 min. The pellet was resuspended in 0.5% (w/v) SDS and again centrifuged at 7,000 *x* g for 10 min. The dilution and centrifugation steps were repeated four times, and the final pellet was washed in 0.5 M NaCl. The pellet was resuspended in ddH_2O, loaded onto a continuous 40-65% sucrose gradient, and centrifuged at 104.000 *x g* for 40 min at 4 ºC. The OB band was collected, diluted 4-fold in ddH_2O, and centrifuged at 7,000 *x g* for 15 min at 4 ºC.

5.2. Electron microscopy

For scanning electron microscopy (SEM), 100 µl of the OB-containing solution (10^9 OBs/ml) were incubated with 300 µl of acetone at 25 ºC for 1 hour. The solution was loaded in a metallic stub, dried overnight at 37 ºC, coated with gold in a Sputter Coater (Balzers) for 3 min, and observed in a scanning electron microscope Jeol JSM 840 A at 10 kV. For transmission electron microscopy (TEM) pellets of purified granules were fixed in Karnovsky fixative (2.5% glutaraldehyde, 2% paraformaldehyde, in 0.1 M, cacodylate buffer, pH 7.2) for 2 h, post-fixed in 1% osmium tetroxide in the same buffer

for 1 h and then stained *en bloc* with 0.5% aqueous uranyl acetate, dehydrated in acetone, and embedded in Spurr's low viscosity embedding medium. The ultrathin sections were contrasted with 2% uranyl acetate and observed in a ZEISS TEM 109 at 80 kV.

5.3. Genomic DNA restriction analyses

Purified granules (10^9 OBs/ml) were dissolved in an alkaline solution and used to extract DNA according to O'Reilly *et al*. (O'Reilly *et al.*, 1992). The quantity and quality of the isolated DNA were determined by electrophoresis on 0.8% agarose (data not shown). The viral DNA (1–2 µg) was individually cleaved with the restriction enzymes *Hin*dIII, *Eco*RI, and *BamH*I (Promega) according to manufacturer's instructions. The DNA fragments generated were analyzed by 0.8% agarose gel electrophoresis (Sambrook & Russel, 2001), visualized, and photographed in AlphaImager® Mini (Alpha Innothech).

5.4. Genome sequencing, assembly, and annotation.

ErelGV genomic DNA was sequenced with the 454 Genome Sequencer (GS) FLX™ Standard (Roche) at the Centro de Genômica de Alto Desempenho do Distrito Federal (Brasília, Brazil). The genome was assembled *de novo* using Geneious 6.0 (Kearse *et al.*, 2012) and confirmed using restriction enzyme digestion profile. The annotation was performed using Geneious 6.0 to identify the open reading frames (ORFs) that started with a methionine codon (ATG) encoding at least 50 amino acids and blastp (33) to identify homologues.

5.5. Phylogeny, genome, and gene comparisons

For *Baculoviridae* phylogenetic analysis, a MAFFT alignment (Katoh *et al.*, 2002) was carried out with concatenated amino acid sequences predicted for 37 baculovirus core genes. A maximum likelihood tree was inferred using PhyML with 100 repetitions of a non parametric bootstrap(Guindon *et al.*, 2010), implemented in Geneious, with LG+I+G+F model selected by Prottest 2.4 (Abascal *et al.*, 2005). Moreover, a genomic comparison was performed using the protein dataset of all the complete *Betabaculovirus* genomes available in Genbank. The dataset was compared using CGView Comparison Tool (Grant *et al.*, 2012)and the results were plotted using CIRCOS (Krzywinski *et al.*, 2009). We also compared ChocGV and PiraGV genomes with ErelGV genome using Mauve alignment (Darling *et al.*, 2004). The horizontal gene transfer (HGTs) events were investigated comparing the maximum likelihood phylogenetic tree inferred using the RAxML method (Stamatakis *et al.*, 2008) and a MAFT alignment of homologues for *he65*-like and *p43*-like,and *Densovirus*-related genes with 100 repetitions of a non parametric bootstrap for branch support.

6. Author's Contributions

Conceived and designed the experiments: DMPAA, FLM, BMR, MLS; Performed the experiments: DMPAA, FLM, MSA, WS; Analyzed the data: DMPAA, FLM, BMR; Contributed reagents/materials/analysis tools: BMR, DMPAA, FLM, MSA, SNB, MLS; Wrote the paper: DMPAA, FLM, BMR, MLS.

7. Acknowledgements

We thank *Conselho Nacional de Desenvolvimento Científico e Tecnológico* (CNPq), *Fundação de Apoio à Pesquisa do Distrito Federal* (FAPDF), *Coordenação de Aperfeiçoamento de Pessoal de Nível Superior* (CAPES), and *EMBRAPA Recursos Genéticos e Biotecnologia* for the financial support; José Osmar Lorenzi for the Picture of *E. ello ello* infected by ErelGV; Ingrid Gracielle Martins da Silva for kindly helping with the sample for microscopy; and Jeffrey J. Hodgson for kindly reviewing the English writing style.

8. References

Abascal, F., Zardoya, R. & Posada, D. (2005). ProtTest: selection of best-fit models of protein evolution. *Bioinformatics* **21**, 2104-2105.

Ackermann, H. & Smirnoff, W. (1983). A morphological investigation of 23 baculoviruses. *J Invertebr Pathol* **41**, 12.

Bellotti, A. C., Arias, B. & Guzman, O. L. (1992). Biological Control of the Cassava Hornworm Erinnyis ello (Lepidoptera: Sphingidae). *The Florida Entomologist* **75**, 10.

Castillo-Acosta, V. M., Aguilar-Pereyra, F., Bart, J. M., Navarro, M., Ruiz-Perez, L. M., Vidal, A. E. & Gonzalez-Pacanowska, D. (2012). Increased uracil insertion in DNA is cytotoxic and increases the frequency of mutation, double strand break formation and VSG switching in Trypanosoma brucei. *DNA Repair (Amst)* **11**, 986-995.

Cui, Q., Shin, W. S., Luo, Y., Tian, J., Cui, H. & Yin, D. (2013). Thymidylate kinase: an old topic brings new perspectives. *Curr Med Chem* **20**, 1286-1305.

D'Amico, V., Slavicek, J., Podgwaite, J. D., Webb, R., Fuester, R. & Peiffer, R. A. (2013). Deletion of v-chiA from a baculovirus reduces horizontal transmission in the field. *Appl Environ Microbiol* **79**, 4056-4064.

Darling, A. C., Mau, B., Blattner, F. R. & Perna, N. T. (2004). Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Res* **14**, 1394-1403.

El-Sharkawy, M. A. (2004). Cassava biology and physiology. *Plant Mol Biol* **56**, 21.

Escasa, S. R., Lauzon, H. A., Mathur, A. C., Krell, P. J. & Arif, B. M. (2006). Sequence analysis of the Choristoneura occidentalis granulovirus genome. *J Gen Virol* **87**, 1917-1933.

Fazolin, M., Estrela, J. L. V., Filho, M. D. C., Santiago, A. C. C. & Frota, F. d. S. (2007). *Manejo Integrado do Mandarová-da-Mandioca Erinnyis ello (L.) (Lepidoptera: Sphingidae): Conceitos e Experiências na Região do Vale do Rio Juruá, Acre*. Rio Branco (AC): EMBRPA Acre (Brazil). Available from: http://iquiri.cpafac.embrapa.br/pdf/doc107.pdf: EMBRAPA Acre.

Ferrelli, M. L., Salvador, R., Biedma, M. E., Berretta, M. F., Haase, S., Sciocco-Cap, A., Ghiringhelli, P. D. & Romanowski, V. (2012). Genome of Epinotia aporema granulovirus

(EpapGV), a polyorganotropic fast killing betabaculovirus with a novel thymidylate kinase gene. *BMC Genomics* **13**, 14.

Finnerty, C. M., Li, G. & Granados, R. R. (2000). Characterization of a granulovirus from the cassava hornworm (Erinnyis ello: Sphingidae). *J Invertebr Pathol* **75**, 273-278.

Garavaglia, M. J., Miele, S. A., Iserte, J. A., Belaich, M. N. & Ghiringhelli, P. D. (2012). The ac53, ac78, ac101, and ac103 genes are newly discovered core genes in the family Baculoviridae. *J Virol* **86**, 12069-12079.

Grant, J. R., Arantes, A. S. & Stothard, P. (2012). Comparing thousands of circular genomes using the CGView Comparison Tool. *BMC Genomics* **13**, 8.

Guillet, M., Van Der Kemp, P. A. & Boiteux, S. (2006). dUTPase activity is critical to maintain genetic stability in Saccharomyces cerevisiae. *Nucleic Acids Res* **34**, 2056-2066.

Guindon, S., Dufayard, J. F., Lefort, V., Anisimova, M., Hordijk, W. & Gascuel, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol* **59**, 307-321.

Harrison, R. L. & Lynn, D. E. (2007). Genomic sequence analysis of a nucleopolyhedrovirus isolated from the diamondback moth, Plutella xylostella. *Virus Genes* **35**, 857-873.

Harrison, R. L. & Popham, H. J. (2008). Genomic sequence analysis of a granulovirus isolated from the Old World bollworm, Helicoverpa armigera. *Virus Genes* **36**, 565-581.

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

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

Herniou, E. A., Arif, B. M., Becnel, J. J., Blissard, G. W., Bonning, B., Harrison, R., Jehle, J. A., Theilmann, D. A. & Vlak, J. M. (2012). Baculoviridae. In *Virus taxonomy Ninth Report of the International Committee on Taxonomy of Viruses* pp. 163-173. Edited by A. M. Q. King, M. J. Adams, E. B. Carstens & E. J. Lefkowitz. San Diego: Elsevier-Academic Press.

Herniou, E. A. & Jehle, J. A. (2007). Baculovirus phylogeny and evolution. *Curr Drug Targets* **8**, 1043-1050.

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

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

Ho, C. K. & Shuman, S. (2002). Bacteriophage T4 RNA ligase 2 (gp24.1) exemplifies a family of RNA ligases found in all phylogenetic domains. *Proc Natl Acad Sci U S A* **99**, 12709-12714.

Hoover, K., Grove, M., Gardner, M., Hughes, D. P., McNeil, J. & Slavicek, J. (2011). A gene for an extended phenotype. *Science* **333**, 1401.

Jehle, J. A., Blissard, G. W., Bonning, B. C., Cory, J. S., Herniou, E. A., Rohrmann, G. F., Theilmann, D. A., Thiem, S. M. & Vlak, J. M. (2006a). On the classification and nomenclature of baculoviruses: a proposal for revision. *Arch Virol* **151**, 1257-1266.

Jehle, J. A., Lange, M., Wang, H., Hu, Z., Wang, Y. & Hauschild, R. (2006b). Molecular identification and phylogenetic analysis of baculoviruses from Lepidoptera. *Virology* **346**, 180-193.

Ji, X., Sutton, G., Evans, G., Axford, D., Owen, R. & Stuart, D. I. (2010). How baculovirus polyhedra fit square pegs into round holes to robustly package viruses. *EMBO J* **29**, 505-514.

Katoh, K., Misawa, K., Kuma, K. & Miyata, T. (2002). MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res* **30**, 3059-3066.

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Meintjes, P. & Drummond, A. (2012). Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28**, 1647-1649.

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

Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., Jones, S. J. & Marra, M. A. (2009). Circos: an information aesthetic for comparative genomics. *Genome Res* **19**, 1639-1645.

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

Lange, M., Wang, H., Zhihong, H. & Jehle, J. A. (2004). Towards a molecular identification and classification system of lepidopteran-specific baculoviruses. *Virology* **325**, 36-47.

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

Lepore, L. S., Roelvink, P. R. & Granados, R. R. (1996). Enhancin, the granulosis virus protein that facilitates nucleopolyhedrovirus (NPV) infections, is a metalloprotease. *J Invertebr Pathol* **68**, 131-140.

Liang, Z., Zhang, X., Yin, X., Cao, S. & Xu, F. (2011). Genomic sequencing and analysis of Clostera anachoreta granulovirus. *Arch Virol* **156**, 1185-1198.

Lima, A. A., Aragao, C. W., de Castro, M. E., Oliveira, J. V., Sosa Gomez, D. R. & Ribeiro, B. M. (2013). A Recombinant MNPV Harboring and Genes from Defective NPV Induce Host Liquefaction and Increased Insecticidal Activity. *PLoS One* **8**, e74592.

Lung, O., Westenberg, M., Vlak, J. M., Zuidema, D. & Blissard, G. W. (2002). Pseudotyping Autographa californica multicapsid nucleopolyhedrovirus (AcMNPV): F proteins from group II NPVs are functionally analogous to AcMNPV GP64. *J Virol* **76**, 5729-5736.

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

Means, J. C. & Passarelli, A. L. (2010). Viral fibroblast growth factor, matrix metalloproteases, and caspases are associated with enhancing systemic infection by baculoviruses. *Proc Natl Acad Sci U S A* **107**, 9825-9830.

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

O'Reilly, D., Miller, L. K. & Luckow, V. A. (1992). *Baculovirus Expression Vectors: a laboratory manual*. New York: Freeman and Company.

Oomens, A. G. & Blissard, G. W. (1999). Requirement for GP64 to drive efficient budding of Autographa californica multicapsid nucleopolyhedrovirus. *Virology* **254**, 297-314.

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

Penades, J. R., Donderis, J., Garcia-Caballer, M., Tormo-Mas, M. A. & Marina, A. (2013). dUTPases, the unexplored family of signalling molecules. *Curr Opin Microbiol* **16**, 163-170.

Pietrowski, V., Ringenberger, R., Rheinheimer, A. R., Bellon, P. P., Gazola, D. & Miranda, A. M. (2010). *Insetos-praga da cultura da mandioca na região Centro-Sul do Brasil*. Marechal Cândido Rondon - Parana: EMBRAPA.

Priet, S., Sire, J. & Querat, G. (2006). Uracils as a cellular weapon against viruses and mechanisms of viral escape. *Curr HIV Res* **4**, 31-42.

Rohrmann, G. F. (2013). *Baculovirus Molecular Biology*. Bethesda (MD): National Center for Biotechnology Information (US). Available from: http://www.ncbi.nlm.nih.gov/books/NBK49500/.

Sambrook, J. & Russel, D. W. (2001). *Molecular Cloning : a laboratory manual*. New York: Cold Spring Harbor.

Schmitt, A. T. (1985). *Eficiência da aplicação de Baculovirus erinnyis no controle do mandarová da mandioca*. Florianópolis, Santa Catarina - Brasil.

Schmitt, A. T. (2002). Principais insetos pragas da mandioca e seu controle. In *Cultura de tuberosas amiláceas Latino Americanas*, p. 7. Edited by M. P. Cereda. Fundação Cargill, São Paulo.

Slack, J. & Arif, B. M. (2007). The baculoviruses occlusion-derived virus: virion structure and function. *Adv Virus Res* **69**, 99-165.

Stamatakis, A., Hoover, P. & Rougemont, J. (2008). A rapid bootstrap algorithm for the RAxML Web servers. *Syst Biol* **57**, 758-771.

Wang, M., Tan, Y., Yin, F., Deng, F., Vlak, J. M., Hu, Z. & Wang, H. (2008). The F protein of Helicoverpa armigera single nucleopolyhedrovirus can be substituted functionally with its homologue from Spodoptera exigua multiple nucleopolyhedrovirus. *J Gen Virol* **89**, 791-798.

Wang, M., Yin, F., Shen, S., Tan, Y., Deng, F., Vlak, J. M., Hu, Z. & Wang, H. (2010a). Partial functional rescue of Helicoverpa armigera single nucleocapsid nucleopolyhedrovirus infectivity by replacement of F protein with GP64 from Autographa californica multicapsid nucleopolyhedrovirus. *J Virol* **84**, 11505-11514.

Wang, R., Deng, F., Hou, D., Zhao, Y., Guo, L., Wang, H. & Hu, Z. (2010b). Proteomics of the Autographa californica nucleopolyhedrovirus budded virions. *J Virol* **84**, 7233-7242.

Wang, Y., Choi, J. Y., Roh, J. Y., Liu, Q., Tao, X. Y., Park, J. B., Kim, J. S. & Je, Y. H. (2011). Genomic sequence analysis of granulovirus isolated from the tobacco cutworm, Spodoptera litura. *PLoS One* **6**, e28163.

Washburn, J. O., Chan, E. Y., Volkman, L. E., Aumiller, J. J. & Jarvis, D. L. (2003). Early synthesis of budded virus envelope fusion protein GP64 enhances Autographa californica multicapsid nucleopolyhedrovirus virulence in orally infected Heliothis virescens. *J Virol* **77**, 280-290.

Wennmann, J. T. & Jehle, J. A. (2014). Detection and quantitation of Agrotis baculoviruses in mixed infections. *J Virol Methods* **197**, 39-46.

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

Yu, M. & Carstens, E. B. (2011). Characterization of an Autographa californica multiple nucleopolyhedrovirus mutant lacking the ac39(p43) gene. *Virus Res* **155**, 300-306.

Zhang, B. Q., Cheng, R. L., Wang, X. F. & Zhang, C. X. (2012). The Genome of Pieris rapae Granulovirus. *J Virol* **86**, 9544.

Zhang, X., Liang, Z., Yin, X., Wang, J. & Shao, X. (2014). Complete genome sequence of Agrotis segetum granulovirus Shanghai strain. *Arch Virol* **http://dx.doi.org/10.1007/s00705-014-2001-y**.

Zhu, S. Y., Yi, J. P., Shen, W. D., Wang, L. Q., He, H. G., Wang, Y., Li, B. & Wang, W. B. (2009). Genomic sequence, organization and characteristics of a new nucleopolyhedrovirus isolated from Clanis bilineata larva. *BMC Genomics* **10**, 91.

9. Supplementary Material

Table S1. Characteristics of the *Erinnyis ello granulovirus* (ErelGV) genome: analysis and homology search. Predicted ORFs are compared with homologous genes in three related genomes.

Orf	Name	Promoter motif	Position	Size (nt)	Size (aa)	CypoGV		ChocGV		PiraGV	
						ORF	Max Id (%)	ORF	Max Id (%)	ORF	Max Id (%)
1[b]	*granulin*	L	1 > 747	747	248	1	96.40	1	97.20	1	87.50
2		L	744 < 1,082	339	112	2	51.60	2	50.00	2	60.30
3[c]	*pk-1*	E	1,063 > 1,899	837	278	3	57.00	3	65.80	3	67.80
4		?	1,996 > 2,541	546	181	-	-	-	-	-	-
5	*dUTPase*-like	E, L	2,811 > 3,764	954	317	16[e]	31.30	-	-	-	-
6		E	3,839 < 4,414	576	191	4	50.30	5	53.40	4	58.50
7		?	4,404 > 4,643	240	79	5	45.20	6	53.30	5	57.90
8[c]	*ie-1*	E	4,743 < 6,059	1,317	438	7	45.10	7	54.70	6	56.30
9[c]		?	6,090 > 6,665	576	191	8	44.70	8	51.50	7	48.40
10[b]		?	6,693 < 6,998	306	101	9	65.30	9	66.30	8	64.40
	hr1		7,113 - 7,204	92							
11*		?	7,151 < 7,792	642	213	-	-	-	-	-	-
12		E	7,791 > 7,949	159	52	-	-	-	-	-	-
13[a]	*odv-e18*	L	8,172 < 8,447	276	91	14	73.60	12	79.50	14	69.40
14[a]	*p49*	E, L	8,448 < 9,827	1,380	459	15	56.20	13	60.60	15	61.10
15*		E	9,736 > 9,960	225	74	-	-	-	-	-	-
16[a]	*odv-e56/pif-5*	L	9,975 < 11,033	1,059	352	18	69.50	14	73.40	16	67.20

17		E, L	11,051	<	11,530	480	159	-	-	-	-	17	29.30
18		E, L	11,545	<	11,943	399	132	-	-	-	-	18	35.40
19		E	12,013	>	12,195	183	60	19	40.70	16	41.80	19	47.30
	hr2		12,183	-	12,222	40							
20	*pep-1*	E, L	12,238	<	12,765	528	175	20	58.90	17	71.30	20	54.00
21	*pep/p10*	E, L	12,885	>	13,910	1,026	341	22	69.20	18	66.40	21	67.60
22[b]	*pep-2*	E, L	13,945	>	14,397	453	150	23	62.20	19	62.80	22	59.90
23		?	14,494	<	15,636	1,143	380	27	24.30	-	-	-	-
	hr3		15,715	-	15,853	139							
	hr4		15,950	-	16,023	74							
24		E	16,087	<	16,248	162	53	-	-	-	-	-	-
25		E	16,439	>	16,723	285	94	-	-	-	-	-	-
26		E	16,647	>	17,879	1,233	410	-	-	22	48.20	24	31.00
27*		E	17,928	<	18,116	189	62	-	-	-	-	-	
28[d]	*f protein*	E	18,356	>	20,137	1,782	593	31	56.60	23	60.90	27	58.70
29		E	20,368	>	21,297	930	309	-	-	-	-	-	-
30		E, L	21,359	<	22,069	711	236	33	31.30	24	44.00	28	43.60
31[a]	*pif-3*	E, L	22,107	>	22,673	567	188	35	52.90	26	47.80	30	54.10
32		E, L	22,695	>	23,006	312	103	39	62.10	28	61.40	31	55.30
33		L	23,038	<	23,352	315	104	40	34.70	-	-	-	-
34[a]	*lef-2*	?	23,501	>	24,028	528	175	41	49.10	29	54.40	33	55.00
35		E	24,012	>	24,284	273	90	42	41.90	30	43.70	34	44.20

36	*he65*-like	L	24,247	<	25,026	780	259	-	-	-	-	-	-
37		E	25,007	<	25,345	339	112	43	29.40	31	38.70	35	37.00
38		E, L	25,355	<	25,801	447	148	45	36.30	32	56.60	36	56.30
39	*mp-nase*	E, L	25,859	<	27,247	1,389	462	46	39.70	33	43.80	37	46.50
40	*p13*	L	27,226	>	28,071	846	281	47	63.40	34	65.50	38	58.70
41	*chtBP*	L	28,091	>	28,348	258	85	9	22.50	7	23.90	8	25.30
42 [a]	*pif-2*	L	28,358	>	29,482	1,125	374	48	70.40	35	69.00	40	69.20
43	*pp-1*	L	29,489	>	29,779	291	96	-	-	36	45.30	-	-
44		L	29,694	>	32,642	2,949	982	50	37.40	-	-	-	-
45 [b]		L	32,639	<	33,292	654	217	52	78.00	37	89.00	43	72.00
46 [c]		L	33,302	>	33,454	153	50	53	63.80	38	70.00	44	57.00
47 [c]	*v-ubq*	E	33,462	<	33,749	288	95	54	82.10	39	87.20	45	85.30
48 [a]		L	33,841	>	34,896	1,056	351	55	52.30	40	66.20	46	63.60
49 [b]		E	34,794	>	35,066	273	90	56	57.40	41	54.30	47	57.40
50 [c]	*39k; pp31*	E	35,067	<	35,903	837	278	57	40.00	42	55.70	48	58.90
51 [b]	*lef-11*	L	35,884	<	36,177	294	97	58	53.30	43	52.20	49	57.40
52	*sod*	L	36,213	<	36,686	474	157	59	65.90	44	70.70	50	68.80
53*		?	36,615	>	36,806	192	63	-	-	-	-	-	45.80
	hr5		36,954	-	37,144	191							
54	*p10*	E, L	36,966	<	37,331	366	121	-	-	45	54.50	-	-
55 [a]	*p74*	E, L	37,344	<	39,308	1,965	654	60	60.10	46	62.70	51	58.80
56		L	39,312	<	39,689	378	125	-	-	53	29.50	-	34.70

57		E	39,763	<	40,482	720	239	-	-	25	35.10	-	-
58		E	40,640	<	41,245	606	201	-	-	48	61.10	54	61.20
59*		L	41,405	<	41,578	174	57	-	-	-	-	-	-
60		L	41,599	<	41,868	270	89	62	76.20	49	79.60	55	71.00
61 [a]	*p47*	E, L	41,943	>	43,124	1,182	393	68	65.50	50	66.70	56	66.30
62 [c]	*bv-e31*	E, L	43,170	>	43,835	666	221	69	66.50	51	67.60	57	67.00
63 [c]	*p24*	L	43,849	>	44,412	564	187	71	62.50	52	67.80	58	67.70
64		?	44,409	>	44,696	288	95	-	-	53	45.50	-	-
65 [c]	*38.7k*	?	44,748	<	45,221	474	157	73	27.40	54	34.80	59	31.50
66 [a]	*lef-1*	?	45,202	<	45,906	705	234	74	57.00	55	64.70	60	63.40
67 [a]	*pif-1*	L	45,937	>	47,547	1,611	536	75	65.50	56	67.20	61	58.70
68	*fgf-1*	?	47,548	<	48,231	684	227	76	45.60	57	55.30	62	54.90
69		E	48,295	<	48,615	321	106	77	37.40	58	50.60	63	44.20
70*		E, L	48,620	>	48,772	153	50	-	-	-	-	-	-
71		E, L	48,785	>	49,291	507	168	79	36.30	59	39.30	64	41.30
72 [c]	*lef-6*	E	49,262	<	49,564	303	100	80	38.00	60	38.80	65	50.00
73 [b]	*dbp*	E	49,649	<	50,494	846	281	81	48.80	61	64.70	66	60.60
74		L	50,509	<	50,739	231	76	82	45.80	62	52.90		
75		E	50,675	<	51,244	570	189	82	31.70	63	38.50	67	41.90
76 [a]	*p48*	E, L	51,268	>	52,440	1,173	390	83	73.20	64	74.80	68	75.10
77 [c]		E, L	52,482	>	52,811	330	109	84	57.40	65	68.40	69	57.40
78 [a]		L	52,865	>	53,989	1,125	374	85	63.50	66	71.30	70	70.20

79[a]	*p6.9*	E, L	54,035	>	54,214	180	59	86	49.10[e]	67	67.80[e]	71	69.40[e]
80[a]	*lef-5*	?	54,264	<	54,995	732	243	87	64.20	68	66.90	72	64.20
81[a]	*38K*	L	54,942	>	55,841	900	299	88	56.10	69	66.00	73	63.50
82[a]	*odv-e28/pif-4*	L	55,838	<	56,323	486	161	89	56.90	70	68.10	74	65.60
83[a]	*dna-helicase-1*	L	56,389	>	59,694	3,306	1101	90	53.00	71	71.30	75	64.30
84[a]	*odv-e25*	E, L	59,724	<	60,365	642	213	91	77.90	72	77.80	76	76.60
85[a]		E, L	60,389	<	60,877	489	162	92	50.90	73	50.90	77	57.50
86[a]	*p33/sox*	L	60,914	>	61,678	765	254	93	63.40	74	66.90	78	68.00
87[a]	*lef-4*	E, L	61,675	<	63,042	1,368	455	95	54.10	75	62.10	80	55.90
88[a]	*vp39*	L	63,119	>	63,979	861	286	96	57.80	76	63.40	81	64.50
89[a]	*odv-e27*	?	64,040	>	64,894	903	300	97	61.50	77	73.10	82	64.20
	hr6		64,964	-	65,177	214							
90*		L	65,057	>	65,233	177	58	-	-	-	-	-	-
91		E, L	65,159	<	66,259	1,101	366	90	40.20	78	45.00	83	41.40
92		?	66,258	>	66,632	375	124	91	48.10	79	40.90	84	41.90
93[a]	*p95/vp91*	E, L	66,619	<	68,409	1,791	596	92	43.00	80	54.50	85	41.80
94[c]		L	68,399	>	68,806	408	135	102	50.00	81	31.70	86	34.10
95[a]		E, L	68,784	>	69,365	582	193	94	67.60	82	72.50	87	67.90
96[a]	*gp41*	E, L	69,343	>	70,179	837	278	95	62.60	83	66.10	88	63.50
97	*iap-3*	E, L	70,213	>	71,022	810	269	17	43.40	84	42.30	79	32.40
98[a]		?	71,032	>	71,322	291	96	105	39.30	85	46.60	89	39.80
99[a]	*vlf-1*	L	71,246	>	72,367	1,122	373	106	68.80	86	65.80	90	70.80

100		E, L	72,429	<	73,097	669	222	-	-	25	25.20	-	-
101		E	73,141	<	73,818	678	225	-	-	87	43.50	-	-
102*		?	73,812	>	73,976	165	54	-	-	-	-	-	-
103		E, L	73,909	>	74,112	204	67	107	59.40	88	71.20	91	67.30
104 [b]		E, L	74,174	>	74,626	453	150	108	58.70	89	64.00	92	63.30
105	*p43*-like	?	74,618	<	75,706	1,089	362	-	-	-	-	-	-
106 [a]	*dna-pol*	E, L	75,743	<	78,901	3,159	1052	111	62.30	90	68.70	93	67.10
107 [a]	*desmoplakin*	?	78,849	>	80,942	2,094	697	112	34.70	91	35.40	94	38.50
108 [c]	*lef-3*	E	81,064	<	82,077	1,014	337	113	41.40	92	60.50	95	52.50
109 [a]	*pif-6*	?	82,049	>	82,426	378	125	114	56.10	93	65.60	96	66.40
110		?	82,477	>	83,064	588	195	115	31.40	94	43.60	97	48.30
111	*iap-5*	E	83,045	>	83,887	843	280	116	48.90	95	56.00	98	53.40
112 [a]	*lef-9*	?	83,865	>	85,346	1,482	493	117	69.40	96	73.40	99	73.60
113 [b]	*fp-25k*	E, L	85,352	>	85,816	465	154	118	65.20	97	70.80	100	63.60
114	*dna-ligase*	?	85,813	<	87,486	1,674	557	120	60.10	99	66.80	102	65.70
115		?	87,658	>	87,888	231	76	121	45.00	100	43.60	103	39.60
116		L	87,943	>	88,161	219	72	122	63.00	101	60.30	104	55.10
117	*fgf-2*	?	88,222	<	89,418	1,197	398	123	34.30	102	34.50	105	34.20
118		E, L	89,545	>	89,823	279	92	124	52.80	103	60.70	106	61.10
119 [a]	*alk-exo*	E	89,873	>	91,078	1,206	401	125	56.30	104	63.90	107	62.50
120	*dna-helicase-2*	E	90,987	>	92,357	1,371	456	126	54.90	105	58.90	108	54.60
121		?	92,402	<	93,457	1056	351	130	41.20	106	41.00	109	32.90

	hr7		92,550	-	92,622	73						
122 [a]	*lef-8*	E, L	93,472	<	96,081	2,610	869	131	70.50	107	73.50	110	71.40
123 [a]		L	96,357	>	96,755	399	132	134	60.20	109	60.20	113	65.40
124		L	96,752	<	97,546	795	264	135	35.60	110	43.20	114	42.90
125	*lef-10*	E	97,773	>	98,159	387	128	137	51.40	112	61.60	115	56.00
126 [a]	*vp1054*	?	98,026	>	99,036	1,011	336	138	57.40	113	63.80	116	64.30
127		L	99,033	>	99,206	174	57	-	-	-	-	117	42.10
128	*fgf-3*	E	99,231	>	100,130	900	299	140	35.90	114	40.20	118	39.80
129	*egt*	E	100,150	<	101,550	1,401	466	141	48.90	115	52.30	119	50.30
130[c]	*me53*	E	101,729	>	102,709	981	326	143	49.00	116	47.00	120	50.20

[a] α-, β-, γ-, and δ-baculovirus core genes; [b] α-, β-, and γ-baculovirus core genes; [c] α- and β-baculovirus core genes; [d] α-, β-, and δ-baculovirus core genes; [e] Identity was achieved by manual alignment. * ErelGV unique genes. The putative gene upstream regions were classified according to the presence of promoter motifs in early (E), late (L), or unknown (?).

Capítulo 4. Characterization of Helicoverpa zea single nucleopolyhedrovirus isolated in Brazil during the first old world bollworm (Noctuidae: *Helicoverpa armigera*) nationwide outbreak

1. Abstract

A baculovirus isolated in Brazil during the first nationwide outbreak of *Helicoverpa armigera* is described by ultrastructural analyses, restriction profiles, pathogenicity of host insects, and complete genome sequence. The results revealed that the virus is an isolate of the species *Helicoverpa zea single nucleopolyhedrovirus* (HzSNPV-Brazilian) never reported before in Brazil. Among the HzSNPV isolates few mutations were observed depicting likely a recent divergence of this lineage. Therefore, the entrance of both foreign pests and natural pathogens into the country must warn the government to reinforce sanitary barriers in order to avoid possible agriculture sabotage and novel foreign pest introductions. Moreover, we found that the Brazilian natural isolate was as lethal as a commercial strain to *H. armigera*. Importantly, virus characterization is of importance in establishment of an economical and useful virus-based biological control program in the country to counteract effectively pest infestations.

Keywords

Helicoverpa argimera, pest outbreak, Brazil, baculovirus, HzSNPV, biological control.

Este capítulo foi inteiramente publicado na revista Virus Review & Research.

Ardisson-Araújo, D. M., Sosa-Gomez, D. R., Melo, F. L., Báo, S. N. & Ribeiro, B. M. (2015). Characterization of Helicoverpa zea single nucleopolyhedrovirus isolated in Brazil during the first old world bollworm (Noctuidae: *Helicoverpa armigera*) nationwide outbreak. *Virus Reviews & Research* **20**, 4.

2. Main Text

In February 2013 the old world cotton bollworm, *Helicoverpa armigera* (Lepidoptera: Noctuidae), that used to be restricted to Africa, Asia, and Europe was identified for the first time in Brazil. A month later, the Brazilian Corporation of Agricultural Research (Portuguese acronym EMBRAPA) reported this occurrence to the Brazilian Ministry of Agriculture, Livestock, and Food Supply (Notification Report n° 70570.000355/2013-2). Unfortunately, by that time the crop pest was already spreadin a high prevalence in the country, which has led to severe agriculture damages and economical losses.This outbreak could be explained by an association of both inadequate management of planting host species (*e.g.* cotton, soybean, and corn) in extensive areas and the uncontrolled use of chemical pesticides which provided together optimal conditions for insect growing.

The genus *Helicoverpa* presents some of the most devastating pest species in the world causing hefty economic losses in several crops including cotton, soybean, wheat, corn, green beans, tomatoes, citrus, and pastures (Cunningham & Zalucki, 2014). The larvae are naturally more tolerant to most of the common insecticides requiring higher application rates to be controlled efficiently (McCaffery, 1998). Almost 30% of all

pesticides used worldwide are directed against *H. armigera* (Ahmad, 2007) although the management of outbreaks has so far been ineffective and also has induced the appearance of resistant insect phenotypes (Oakeshott *et al.*, 2013; Rowley *et al.*, 2011) including engineered plants expressing *Bacillus thuringiensis* (Bt) toxins (Tabashnik *et al.*, 2009).Therefore, other naturally found disease-causing pathogens like baculoviruses are important alternatives for the integrated and effective control of *Helicoverpa* (Rowley *et al.*, 2011). Robust virus characterization allows the establishment of a virus-based biological control program to control pest outbreaks as a safety, useful, and economical alternative for chemical pesticides.

For the crop season 2013/2014, commercial baculoviruses infective to the old world bollworm have been imported to be usedin Brazil. Before this allowance by the Brazilian government to import *Helicoverpa*-infecting baculoviruses from other countries in order to control a nationwide *H. armigera* outbreak, a baculovirus was isolated in field from larvae cadavers with symptoms of infection. Cadavers of *H. armigera* were collected in March/2013 on soybean crops in Warta, Londrina County, Parana, Brazil.Although *H. zea* does not infest soybean in Brazil, we confirmed the species *H. armigera* by amplifying and sequencing the genes *cytochrome c oxidase I* (COI), *cytochrome B*, and the region *cox1-tRNA-leu-cox2* (data not shown). Electron microscopy (EM) of purified occlusion bodies (OBs), which are hallmarks of the family *Baculoviridae*, showed polyhedral shape (FIG. 1A) and virions with singly enveloped nucleocapsids within (FIG. 1B). The occlusion bodies purification, polyhedra EM and DNA extraction were performed according to published protocols (Ardisson-Araujo *et al.*, 2014). The viral DNA (1–2 µg) was individually cleaved with the restriction enzymes *Xho*I, *Bgl*II, *Pst*I, or *Bam*HI (Promega) according to manufacturer's

instructions. Importantly, HzSNPV is found naturally infecting the genus *Helicoverpa* during its larval stage (Chen *et al.*, 2002; Ogembo *et al.*, 2009; Rowley *et al.*, 2011). Based on the comparison of both the viral DNA restriction enzyme profiles (FIG. 1C) and previously published data of other *Helicoverpa*-infecting nucleopolyhedroviruses (Chen *et al.*, 2002), we concluded that the virus belonged to the species HzSNPV which was one of the first commercial baculovirus pesticides registered in the 1970's (Virion-H, Biocontrol-VHZ, Elcar) and has been so far produced and applied successfully against both *H. armigera* and *H. zea* (Rowley *et al.*, 2011; Shieh, 1989; van Beek & Davis, 2007). Therefore, we named the Brazilian isolate HzSNPV-Brazilian, even being found in *H. armigera* cadavers.

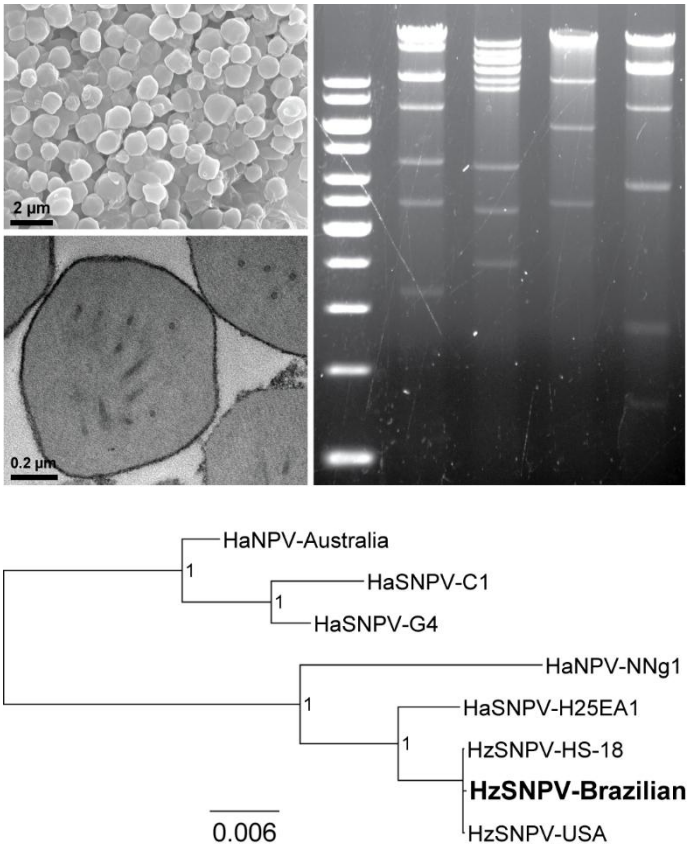

FIG. 1. Characterization of the *Helicoverpa*-infecting baculovirus found in Brazil. (A) Scanning electron micrograph shows polyhedral-shaped OBs. (B) Transmission electron micrograph shows sliced OBs with single-enveloped nucleocapsids within. (C) Agarose gel electrophoresis-resolved HzSNPV-Brazilian genome DNA fragments digested with *Xho*I (lane 1), *Bgl*II (lane 2), *Pst*I (lane 3), and *Bam*HI (lane 4), and molecular weight marker (lane M). All the features together corroborate that this isolate belongs to the species *Helicoverpa zea single nucleopolyhedrovirus* (HzSNPV). (D) Maximum likelihood tree of *Helicoverpa*-isolated single nucleopolyhedroviruses. The phylogeny was inferred using MAFFT alignment of whole genome and the relationship using PhyML method. The Brazilian isolate (boldface) is related to both HzSNPV-USA and HzSNPV-HS18 viruses and the closest relative to this group is the Australian HaNPV-H25EA1. Branch support is estimated by a Shimodaira–Hasegawa-like test.

To further substantiate our data, we carried out a bioassay using the Brazilian strain and a commercially available virus from the same species HzSNPV (Gemstar®) towards *H. armigera* and *H. zea*. For this experiment, serial dilution of the virus were carried out to determine both LC_{50} and LC_{99} in third-instar caterpillars and mixed with the larva diet as previously described (Ardisson-Araujo *et al.*, 2014). Insects were allowed to feed *ad libitum* on virus inoculated diet. A group with no treatment (n=60) was set up as control. Mortality was recorded 13 days post-infection (p.i.) by scoring the number of dead insect which had no response to touch. The data was analyzed by Polo Plus program (LEORA SOFTWARE, POLO-Plus 1.0, Probit and Logit analysis, Petaluma, California. 2003). We found that the Brazilian isolate virus was more lethal to *H. zea* than to *H. armigera* in oral bioassays (Table 1). The OB concentration per ml of artificial diet capable to kill 50% of the tested insects at the third-instar (LC50) was 987

OB/ml to *H. armigera* and 215 OB/ml to *H. zea*. This ability to kill *H. zea* more efficiently by HzSNPV was previously reported (Rowley *et al.*, 2011), which is a very interesting aspect of short term adaptation to the host even presenting high identity to the closest relatives (*i.e.*HaNPV isolates). Moreover, we tested whether the Brazilian strain could be as efficient as the commercially available HzSNPV from Gemstar® (Certis, Columbia, USA) to kill *H. armigera*. We found that both viruses had statistically equal lethal concentration to the tested insect (Table 1). Conversely, in a worldwide *Helicoverpa*-isolated baculovirus study, Gemstar® isolate of HzSNPV presented lethal concentration higher than the other naturally found isolates (Rowley *et al.*, 2011).

Table 1. Dose-mortality responses of *Helicoverpa* spp. third instar larvae infected orally with either HzSNPV-Brazilian (Br) or a commercial strain of HzSNPV (Gemstar®).

Insect	Virus	n[1]	LC_{50} (OB/ml)	95% Fiducial limits		LC_{99} (OB/ml)
				Lower	Upper	
H. zea	Br	197	2.15×10^2	0.75×10^2	4.00×10^2	130.0×10^2
H. armigera	Br	482	9.87×10^2	6.60×10^2	15.6×10^2	754.0×10^2
	Gemstar®	283	10.2×10^2	4.71×10^2	21.5×10^2	nt

[1], number of tested insects; nt, non-tested

The whole genome of HzSNPV-Brazilian (Genbank: KM596835) was sequenced with the 454 Genome Sequencer (GS) FLX™ Standard (Roche) at the Center of High-performance Genomic (Brasilia, Brazil). The genome was *de novo* assembled using Geneious 6.0 (Kearse *et al.*, 2012) and confirmed with the digestion profile. Annotation was also performed using Geneious 6.0 to identify the open reading frames (ORFs) that started with a methionine codon (ATG) encoding polypeptides with at least 50 amino acids, and BLASTP (Altschul *et al.*, 1997) to identify homologs. The sequencing produced 8,237 single-end reads. After size and quality trimming, 8,068 reads (average

size of 755.5 nt) were assembled with coverage of 47.2±12.0 bp/site. The HzSNPV-Brazilian genome has a size of 129,694 bp with a G+C content of 39.1 %. The genome potentially codes for 146 putative ORFs with predicted polypeptides of at least 50 amino acids and all of them are homologs to those of HzSNPV isolates. Eight ORFs were not annotated in the first described genome but were present. Isolates of HzSNPV have a nucleotide pairwise alignment identity of 99% and the average identity across the *Helicoverpa*-infecting SNPVs is 96.22±1.49%. HzSNPV-Brazilian presents a deletion of 1,000 bp in the homolog region 1 (confirmed by PCR, data not shown).

For phylogenetic analysis, a MAFFT alignment (Katoh *et al.*, 2002)was carried out with whole genome sequences of all *Helicoverpa*-isolate single nucleopolyhedrovirus available in Genbank. This alignment was manually inspected, and poorly aligned regions (at least 50 % of gaps) were deleted. The resulting alignment was approximately 135 kb long. The maximum likelihood tree was inferred using PhyML(Guindon *et al.*, 2010), under Tamura-Nei model selected by jModelTest-2.1.4 software (Darriba *et al.*, 2012). The branch support was estimated by a Shimodaira-Hasegawa-like test (Anisimova *et al.*, 2011). The phylogenetic analysis confirmed that HzSNPV-Brazilian is closely related to HzSNPV isolates (FIG. 1D).The short branch length compared to the other isolatesindicates low genetic diversity and low branch support prevented us to establish the origin of the Brazilian strain.

In order to determine the CDS diversity among the *Helicoverpa*-infecting single nucleopolyhedrovirus, we considered the completely sequenced viruses as two separated groups including viruses isolated from (*i*) *H. armigera* and (*ii*) from *H. zea*. To search for polymorphism, we concatenated 135 ORFs found to be common among

all the *Helicoverpa*-isolated single nucleopolyhedrovirus: HaNPV isolates C1 [AF303045], Australia [JN584482], G4 [AF271059], NNg1 [AP010907], and H25A1 [KJ922128] and HzSNPV isolates USA [AF334030], HS18 [KJ004000], and Brazilian [KM596835]. We performed a MAFFT alignment and set as reference sequence the genome of the G4 for the HaNPVgroup and the Brazilian for HzSNPV group.For the first (*i.e.* HaNPV-related baculovirus), we found 624 nonsynonymous polymorphisms out of 1,592 (data not shown). On the other hand, we found only 13 non synonymous polymorphisms out of 15 among the three HzSNPV isolates (data not shown). This very low genetic diversity among the HzSNPV isolates in comparison to HaNPV depicts a recent divergence of the isolates reinforcing the hypothesis that the Brazilian isolate could be recently introduced into the country from either the American or the Russian strain. Sublethal and latent infections are of importance for the persistence of baculoviruses in the environment (Kukan, 1999) which could explain how HzSNPV together with the host insect has gotten into the country. In a previous study, we found the first non-Asian isolate of a *Bomby mori*-infecting baculovirus in Brazil. By complete genome sequencing and phylogenetic analysis, similarly to the results found in this work, we found that the virus was probably introduced together with the insect into the country (Ardisson-Araujo *et al.*, 2014).

We determined the following from the present short report. (*i*) The *H. armigera*-infecting baculovirus isolated in Brazil belongs to the species HzSNPV. (*ii*) It is a single NPV with polyhedral-shaped occlusion bodies. (*iii*) The virus was more lethal to *H. zea* than to *H. armigera*, besides of presenting the same lethality as that observed for the commercial strain Gemstar® to *H. armigera*. (*iv*) The complete genome sequence

revealed its close relationship to HzSNPV isolates. (*v*) Low genetic diversity was observed among the HzSNPV isolates.

3. Acknowledgements

Wethank 'Conselho Nacional de Desenvolvimento Científico e Tecnológico' (CNPq), 'Fundação de Apoio à Pesquisa do Distrito Federal' (FAPDF), and 'Coordenação de Aperfeiçoamento de Pessoal de Nível Superior' (CAPES) for financial support; Ingrid Gracielle Martins da Silva for kindlyhelpingwiththesample for TEM and SEM.

4. References

Ahmad, M. (2007). Insecticide resistance mechanisms and their management in Helicoverpa armigera (Hubner) – A review. *Journal of Agricultural Research* **45**, 17.

Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25**, 3389-3402.

Anisimova, M., Gil, M., Dufayard, J. F., Dessimoz, C. & Gascuel, O. (2011). Survey of branch support methods demonstrates accuracy, power, and robustness of fast likelihood-based approximation schemes. *Syst Biol* **60**, 685-699.

Ardisson-Araujo, D. M., Melo, F. L., de Souza Andrade, M., Brancalhao, R. M., Bao, S. N. & Ribeiro, B. M. (2014). Complete genome sequence of the first non-Asian isolate of Bombyx mori nucleopolyhedrovirus. *Virus Genes*.

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

Cunningham, J. P. & Zalucki, M. P. (2014). Understanding heliothine (Lepidoptera: Heliothinae) pests: what is a host plant? *J Econ Entomol* **107**, 881-896.

Darriba, D., Taboada, G. L., Doallo, R. & Posada, D. (2012). jModelTest 2: more models, new heuristics and parallel computing. *Nat Methods* **9**, 772.

Guindon, S., Dufayard, J. F., Lefort, V., Anisimova, M., Hordijk, W. & Gascuel, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol* **59**, 307-321.

Katoh, K., Misawa, K., Kuma, K. & Miyata, T. (2002). MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res* **30**, 3059-3066.

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Meintjes, P. & Drummond, A. (2012). Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28**, 1647-1649.

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

McCaffery, A. R. (1998). Resistance to insecticides in Heliothine Lepidoptera: a global view. *Philosophical Transactions of the Royal Society B: Biological Sciences* **353**, 16.

Oakeshott, J. G., Farnsworth, C. A., East, P. D., Scott, C., Han, Y., Wu, Y. & Russell, R. J. (2013). How many genetic options for evolving insecticide resistance in heliothine and spodopteran pests? *Pest Manag Sci* **69**, 889-896.

Ogembo, J. G., Caoili, B. L., Shikata, M., Chaeychomsri, S., Kobayashi, M. & Ikeda, M. (2009). Comparative genomic sequence analysis of novel Helicoverpa armigera nucleopolyhedrovirus (NPV) isolated from Kenya and three other previously sequenced Helicoverpa spp. NPVs. *Virus Genes* **39**, 261-272.

Rowley, D. L., Popham, H. J. & Harrison, R. L. (2011). Genetic variation and virulence of nucleopolyhedroviruses isolated worldwide from the heliothine pests Helicoverpa armigera, Helicoverpa zea, and Heliothis virescens. *J Invertebr Pathol* **107**, 112-126.

Shieh, T. R. (1989). Industrial production of viral pesticides. *Adv Virus Res* **36**, 315-343.

Tabashnik, B. E., Van Rensburg, J. B. & Carriere, Y. (2009). Field-evolved insect resistance to Bt crops: definition, theory, and data. *J Econ Entomol* **102**, 2011-2025.

van Beek, N. & Davis, D. C. (2007). Baculovirus insecticide production in insect larvae. *Methods Mol Biol* **388**, 367-378.

Capítulo 5. Functional characterization of *hesp018*, a baculovirus-encoded serpin gene

1. Summary

The serpin family of serine proteinase inhibitors plays key roles in a variety of biochemical pathways. In insects, one of the important functions carried out by serpins is regulation of the phenoloxidase cascade, a pathway that produces melanin and other compounds that are important in insect humoral immunity. Recent sequencing of the baculovirus Hemileuca sp. nucleopolyhedrovirus (HespNPV) genome revealed the presence of a gene, *hesp018*, with homology to insect serpins. To our knowledge *hesp018* is the only intact serpin homolog known to exist in a viral genome outside of the chordopoxviruses. In this study, the Hesp018 protein was shown to be a functional serpin with inhibitory activity against a subset of serine proteinases. Hesp018 also inhibited phenoloxidase activation when mixed with lepidopteran hemolymph. The Protein was secreted when expressed in lepidopteran cells, and a baculovirus expressing it exhibited accelerated production of viral progeny during *in vitro* infection. Expression of Hesp018 also reduced caspase activity induced by baculovirus infection, but caused increased cathepsin activity. In infected insect larvae, expression of Hesp018 resulted in faster larval melanization, consistent with increased activity of viral cathepsin. Finally, expression of Hesp018 increased the virulence of a prototype baculovirus by 4-fold in orally-infected neonate *Trichoplusia ni* larvae. Based on our observations, we hypothesize that the *hesp018* may have been retained in HespNPV due to its ability to inhibit the activity of select host proteinases, possibly including proteinases involved in the phenoloxidase response, during infection of host insects.

Este capítulo foi inteiramente publicado na revista Journal of General Virology. **Ardisson-Araujo, D. M., Rohrmann, G. F., Ribeiro, B. M. & Clem, R. J. (2015).** Functional characterization of hesp018, a baculovirus-encoded serpin gene. *J Gen Virol* **96**, 1150-1160.

2. Introduction

The insect innate immune system responds against invading pathogens and parasites (Jiang *et al.*, 2010; Xu & Cherry, 2014) by means of antimicrobial peptides, the action of hemocytes, and by intracellular mechanisms such as RNA interference (Jayachandran *et al.*, 2012) and apoptosis (Clem, 2005). On the other hand, infectious agents have evolved mechanisms to overcome or even manipulate this hostile environment in order to survive and reproduce (Clem & Passarelli, 2013; Ferrandon *et al.*, 2007).

Baculoviruses are large DNA viruses that mainly infect the larval stages of Lepidoptera (moths and butterflies) (Herniou *et al.*, 2012; Rohrmann, 2013a). A typical baculovirus infection initiates when susceptible caterpillars feed on foliage contaminated with viral occlusion bodies (OBs), which release occlusion-derived virions in the midgut and establish primary infection (Slack & Arif, 2007). Infected midgut epithelial cells produce budded virions (BV), which cross the midgut barrier and cause systemic secondary infection. Baculoviruses are able to manipulate the cellular environment to enhance their infection (Thiem, 2009), for example by inhibiting cell cycle progression (Prikhod'ko & Miller, 1998), inducing DNA damage response (Huang *et al.*, 2011), blocking apoptosis (Ikeda *et al.*, 2011), and inducing shutoff of host gene expression

(Ooi & Miller, 1988). They are also able to manipulate host physiology and behavior through the expression of various viral proteins (Kamita *et al.*, 2005; Katsuma *et al.*, 2012; O'Reilly & Miller, 1989). There is evidence that baculovirus infection causes immune suppression in infected larvae, which leads to an increase in gut microbiota (Jakubowska *et al.*, 2013). However, it is not clear whether baculoviruses can directly (*i.e.* by expression of a viral gene product) control humoral innate immune responses.

The presence of bacteria or fungi in the lepidopteran hemocoel stimulates cellular and humoral responses (Jiang *et al.*, 2010); however, there is no consensus regarding immune activation caused by baculovirus infection. It has been proposed to be dependent on the route of infection or on specific responses which might vary between insects (Terenius *et al.*, 2009). Hemocytes, when induced, can neutralize pathogens by engulfing or trapping them in nodules (Dean *et al.*, 2004; Yu & Kanost, 2004) which become melanized through the activation of phenoloxidases (POs). POs produce reactive intermediates for melanin production and these contribute to the killing of microbes (Nappi & Christensen, 2005; Zhao *et al.*, 2007; Zhao *et al.*, 2011). Baculovirus-infected cells can also be encapsulated by hemocytes and melanized (McNeil *et al.*, 2010; Trudeau *et al.*, 2001; Washburn *et al.*, 2000). POs are present in insect plasma in an inactive form called proPO. Microorganism invasion triggers activation of a serine proteinase cascade, which eventually results in the cleavage of proPO to active PO. The activity of the proteinases in the PO cascade is negatively regulated by serine proteinase inhibitors (serpins) (Jiang *et al.*, 2010).

In addition to regulating the activation of PO, serpins are also important in many other pathways involving serine proteinases (Gettins, 2002; Silverman *et al.*, 2001). Serpins

consist of a single peptide chain typically composed of three β-sheets (A, B and C) and several α-helices. A reactive center loop located between β-sheet A and C determines the inhibitory selectivity (Huntington, 2011). Serpin inhibition initially involves formation of a non-covalent complex with the targeted proteinase. When the enzyme cleaves the serpin at the P1 residue within the reactive center loop, a covalent ester linkage is formed between the serpin and proteinase, resulting in dramatic conformational changes in both the enzyme, which is now inactive, and the serpin, which is cleaved (Huntington, 2011).

While serpin homologs are present in chordopoxviruses that infect vertebrates, where they inhibit apoptosis and inflammatory responses (Tewari *et al.*, 1995), until recently intact serpin genes had not been reported in other virus genomes. However Rohrmann *et al*.(Rohrmann *et al.*, 2013) recently reported the presence of a serpin homolog, *hesp018*, in the genome of the baculovirus *Hemileuca sp.* nucleopolyhedrovirus (HespNPV). In this report, we describe the analysis of the phylogeny of this baculovirus-encoded gene and its ability to function as a serpin, as well as the effects of hesp018 expression on the fitness of a prototype baculovirus.

3. Results

3.1. Phylogenetic analysis of the *hesp018* gene

To investigate the relationship of Hesp018 to other serpins, a maximum likelihood tree was constructed using the predicted amino acid sequence of Hesp018 and several arthropod serpin sequences. The results support the hypothesis that *hesp018* arose as a

horizontal gene transfer from a lepidopteran host (Fig. 1A). We found that the Hesp018 sequence clustered with lepidopteran serpin type 4 orthologs, as previously hypothesized (Rohrmann *et al.*, 2013). Interestingly, the region of the HespNPV genome containing *hesp018* may have been a hotspot for recombination in the ancestor of HespNPV-related baculovirus species (Group II *Alphabaculovirus* species) since hypothetical newly acquired genes and repeat regions are found in this region (Fig.1B).

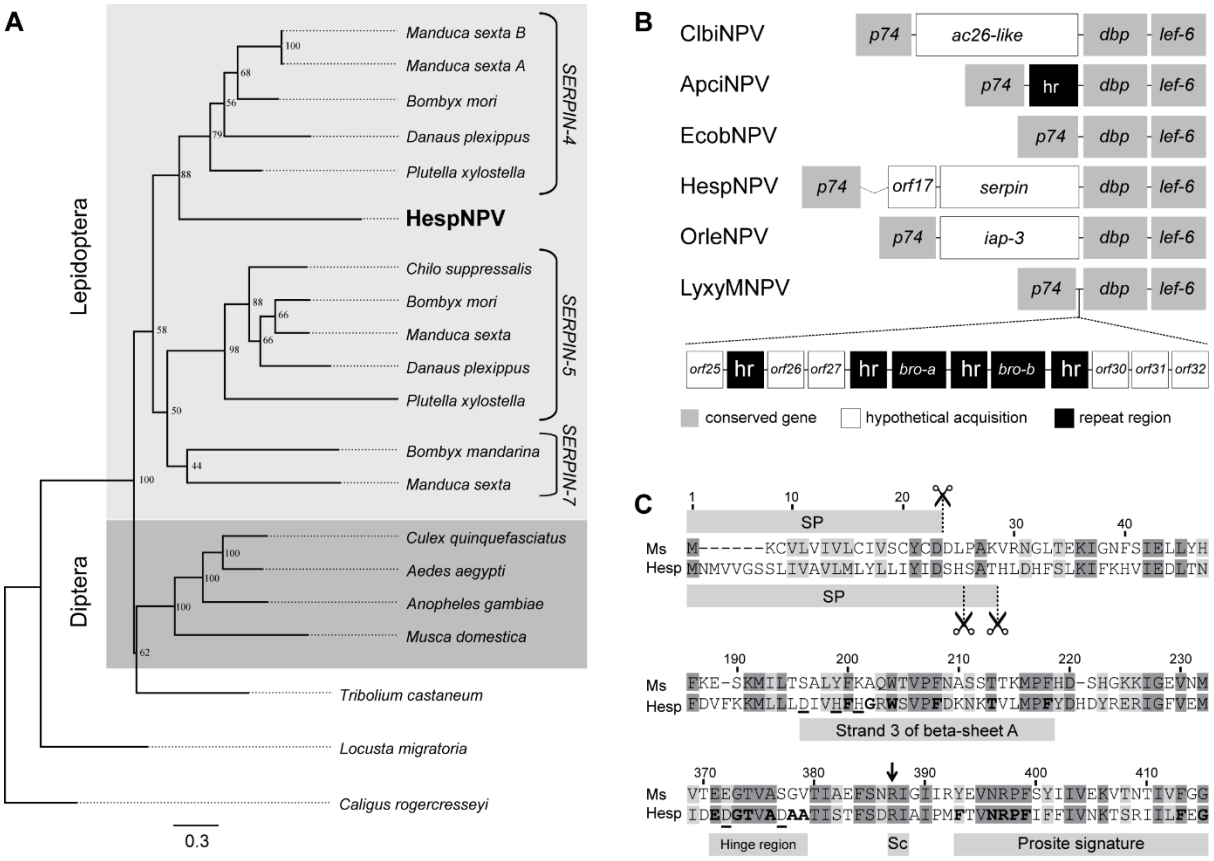

FIG. 1. *In silico* analyses of Hesp018. A) Phylogenetic analysis of selected arthropod serpins. Hesp018 sequence clustered with lepidopteran type 4 serpins. A crustacean-derived serpin roots the Maximum Likelihood tree. B) The *hesp018* gene region is a possible hotspot for recombination events. Gene order is shown from *Hemileuca sp.* nucleopolyhedrovirus (HespNPV) and other type II alphabaculoviruses including *Clanis*

bilineata NPV (ClbiNPV), *Apocheima cinerarium* NPV (ApciNPV), *Ectropis obliqua* NPV (EcobNPV), *Orgyia leucostigma* NPV (OrleNPV), and *Lymantria xylina* MNPV (LyxyMNPV). C) Serpin annotation based on alignment of *M. sexta* and Hesp018 protein sequences. SP, signal peptide with predicted cleavage sites represented by scissors and dashed lines; Sc, scissile bond involved in inhibition by serpins. The arginine residue at the predicted P1 site is indicated by an arrow. Conserved residues are bold-faced and residues that are in the signature but are not conserved are underlined.

Based on alignment with insect serpins, we found that Hesp018 contains a predicted signal peptide, a conserved strand 3 of beta-sheet A, a hinge region, a scissile bond, and a prosite signature, all features commonly found in serpin proteins (Fig. 1C). The signal peptide contains two potential predicted cleavage sites (Fig. 1C, scissors). For the serpin signature region, six of nine residues are conserved in strand 3 of beta-sheet A, six of eight are conserved in the hinge region and the prosite-signature contains seven conserved residues (Fig. 1C). Importantly, Hesp018 has a basic arginine residue at the predicted P1 site at the scissile bond region (Fig. 1C, arrow), characteristic of trypsin-like serine proteinase inhibitors (An *et al.*, 2012). Based on these characteristics, we predicted that Hesp018 is an active serpin.

3.2. Inhibitory activity of the baculovirus serpin

To test for serpin activity, His-tagged Hesp018 protein was expressed in *E. coli*, purified and incubated with the serine proteinases trypsin, chymotrypsin, plasmin, and proteinase K. Hesp018 efficiently inhibited in a concentration-dependent manner trypsin, chymotrypsin, and plasmin, but not proteinase K (Fig. 2 and Table 1). Trypsin

and chymotrypsin were each able to cleave Hesp018 (Fig. 2E). Of the enzymes tested, plasmin was the most sensitive to inhibition by Hesp018, reaching 100% inhibition at 10:1 molar ratio (serpin:proteinase). Inhibition of trypsin was 87% at 10:1, while chymotrypsin reached 67% inhibition at 10:1 (Table 1). Plasmin formed a stable complex with Hesp018 as detected by immunoblot analysis (data not shown), indicating that Hesp018 inhibited plasmin by the conserved serpin mechanism.

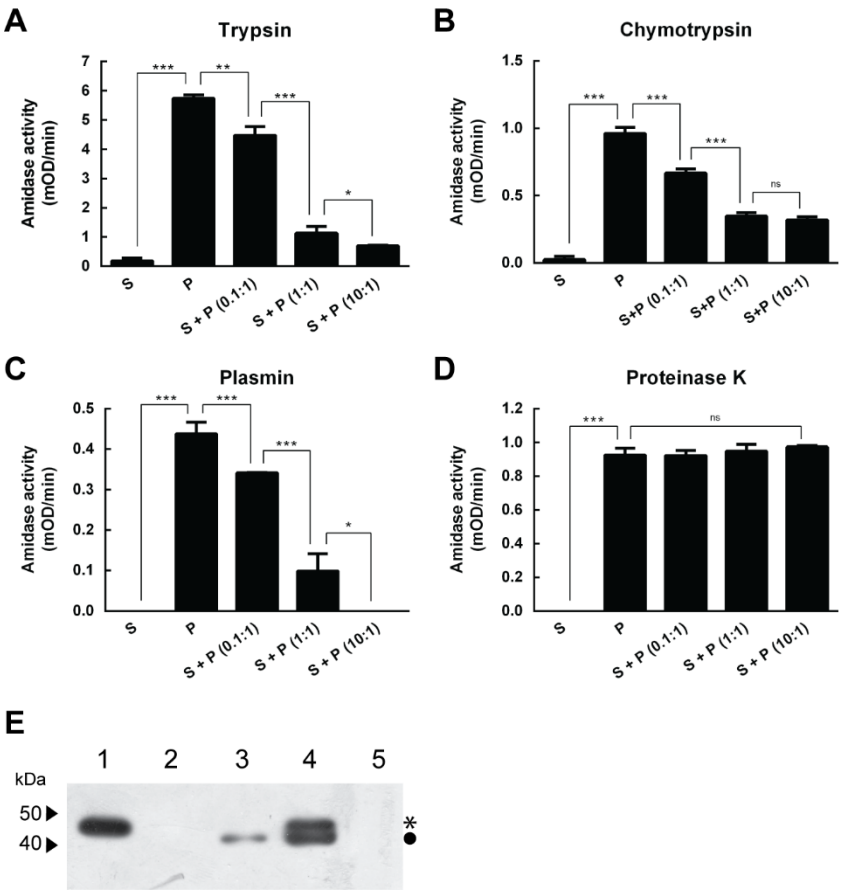

FIG. 2. Hesp018 inhibits a subset of serine proteinases. Recombinant Hesp018 protein was incubated with (A) trypsin, (B) chymotrypsin, (C) plasmin, or (D) proteinase K at the indicated molar ratios of serpin (S) and proteinase (P), after which the residual amidase activity was measured. Standard errors (n=3) and statistical differences obtained by unpaired two-tailed Student's t test are indicated (*p* values: *, $p<0.05$; **,

$p \leq 0.01$; ***, $p \leq 0.0001$). (E) The mixtures for trypsin and chymotrypsin were subjected to SDS-PAGE under reducing conditions, and the cleaved serpin was detected by immunoblotting using anti-His antibody: Lane 1, Hesp018; lane 2, trypsin; lane 3, Hesp018 + trypsin (1:1); lane 4, Hesp018 + chymotrypsin (1:1); lane 5, chymotrypsin. The asterisk and the circle indicate the migration of full length and cleaved Hesp018, respectively.

Table 1. Inhibition of serine proteinases by Hesp018.

Ratio (Serpin:Proteinase)	Trypsin	Chymotrypsin	Plasmin	Proteinase K
0.1:1	22.0[a] ± 5.1	30.6 ± 3.2	22.0 ± 0.3	0.3 ± 3.3
1:1	80.2 ± 4.0	63.8 ± 2.8	79.1 ± 9.3	-2.5 ± 4.4
10:1	87.8 ± 0.4	66.7 ± 2.4	100.0 ± 0.0	-5.3 ± 0.9

[a]Percent inhibition relative to control lacking Hesp018 (± SE).

Serpin-4 from *Manduca sexta* has been shown to inhibit the hemolymph proteinases HP-1, HP-6, and HP-21 in the PO pathway (Tong *et al.*, 2005). Since the closest relatives of Hesp018 are serpin-4 homologs, we examined the ability of purified recombinant Hesp018 protein to inhibit PO activation in lepidopteran plasma. We found that Hesp018 was able to prevent bacteria-stimulated *M. sexta* plasma PO activity (Fig. 3A) and migrated faster by SDS-PAGE after incubation with the insect plasma (Fig. 3B), indicating that Hesp018 was cleaved and functioned as a substrate inhibitor in hemolymph. Together, the results in Figs. 2 and 3 indicate that Hesp018 is a functional serpin.

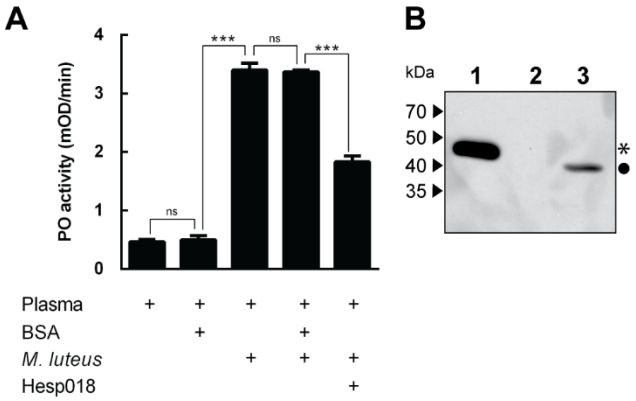

FIG. 3. Hesp018 can inhibit PO activity and is cleaved in plasma. (A) Inhibition of ProPO activation in *M. sexta* plasma by Hesp018. Cell-free plasma was incubated with 100 ng purified Hesp018 for 10 min, followed by addition of *M. luteus* extract to stimulate PO activation. PO activity (mean ± S.E., n = 3) was measured after 10 min. Asterisks indicate statistical difference obtained by unpaired two-tailed Student's t test ($p \leq 0.0001$) and by one-way ANOVA (Graphpad) ($p < 0.0001$). (B) Purified Hesp018 was incubated with plasma for 10 min and the mixture activated with bacterial extract. Transferred proteins were then detected with anti-His antibody; (1) Hesp018, (2) plasma, and (3) plasma-treated Hesp018. The asterisk and the circle indicate the migration of full length and cleaved Hesp018, respectively.

3.3. Serpin expression accelerates AcMPNV BV production

While it would be ideal to test the function of Hesp018 in the context of HespNPV infection, this virus is biologically uncharacterized, and currently only exists as an archived sample of occlusion bodies (Rohrmann, 2013b; Rohrmann *et al.*, 2013). Therefore to examine the effect of Hesp018 expression during baculovirus infection, we constructed recombinant versions of the prototype Group I alphabaculovirus,

Autographa californica M nucleopolyhedrovirus (AcMNPV), expressing Hesp018 with or without a C-terminal HA epitope tag (Fig. 4A). For comparison we also constructed a virus expressing *M. sexta* serpin-4B protein (Fig. 4A).

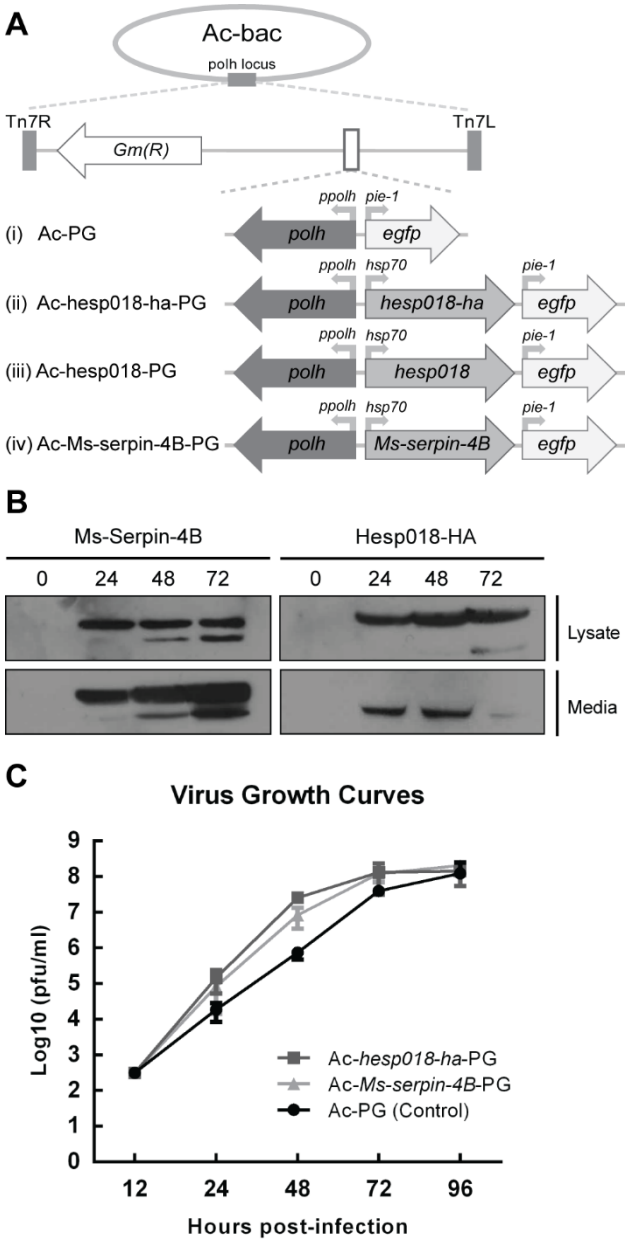

FIG. 4. Recombinant versions of AcMNPV expressing serpin genes. (A) Schematic representation of viruses expressing either Hesp018 or *M. sexta* serpin-4B under control of the *D. melanogaster* hsp70 promoter. (B) Hesp018 and serpin-4B were secreted

during recombinant AcMNPV *in vitro* infection. Cells were harvested at the indicated times and cell lysates or supernatants were analyzed by immunoblotting with either anti-HA antibody or anti-*M. sexta* serpin-4B. (C) BV growth curves (MOI = 0.01) as determined by $TCID_{50}$ assay. The growth curve for the tagged *hesp018*-expressing virus was similar to the untagged version (not shown). Expression of either Hesp018 or serpin-4B increased BV titers at 24 and 48 h p.i. (n=4) with significance levels of $p<0.001$ obtained by Student's t test.

When recombinant viruses expressing Hesp018-HA or serpin-4B were used to infect Sf9 cells, both proteins were secreted into the media (Fig.4B). Expression of the *serpin* genes resulted in significantly increased AcMNPV BV titers at 24 and 48 h p.i., although by 96 h p.i. the titers were similar to control virus (Fig. 4C). This result indicates that serpin expression accelerated the production of BV *in vitro*.

3.4. Viral and cellular enzyme activities influenced by Hesp018 expression

We next tested whether the *hesp018* gene could affect, directly or indirectly, cellular or viral proteinase activities when expressed during recombinant AcMNPV infection *in vitro*. Most baculoviruses, including AcMNPV, encode a papain-like cysteine proteinase that has homology to cathepsins (v-cath), as well as a chitinase homolog (chiA); both v-cath and chiA are required for baculovirus-induced host liquefaction, and v-cath is also involved in viral-induced host melanization (Hawtin *et al.*, 1997; Slack *et al.*, 1995). In addition, AcMNPV infection stimulates the activation of cellular effector caspases, whose activities are normally inhibited by the viral P35 protein but which can be studied using mutants lacking *p35*.

Interestingly, we found that infection of Sf9 cells with viruses expressing either baculovirus- or insect-derived *serpin* genes caused significantly increased cathepsin activity, but not chitinase activity, when compared to the parental virus Ac-PG (Fig. 5A and B). Since cathepsin activity is required for caterpillar melanization (Slack *et al.*, 1995), we also examined the timing of melanization of *M. sexta* larvae injected with BV. There was a noticeable increase in melanization at 24 h post-death when the larvae were injected with viruses expressing either Hesp018 or serpin-4B, compared to control virus (Fig. 5C). This increased melanization suggests that cathepsin activity may have also been increased by serpin expression during *in vivo* infection.

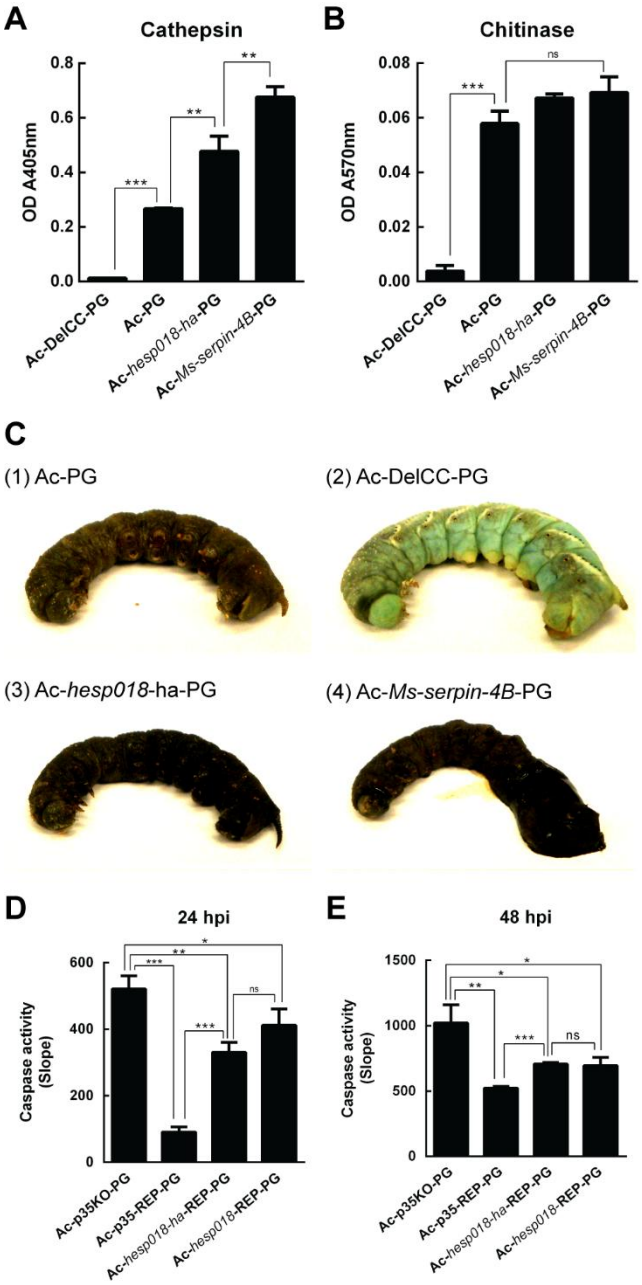

FIG. 5. Viral and cellular enzyme activities influenced by recombinant baculovirus-mediated Hesp018 expression. Sf9 cells were infected with the indicated viruses and A) cathepsin or B) chitinase activity was measured at 42 h p.i. Ac-DelCC-PG is a virus lacking both chitinase and cathepsin viral genes. C) Expression of Hesp018 or serpin-4B accelerates post-mortem melanization of *M. sexta* larvae. Larvae infected with viruses expressing Hesp018 or serpin-4B had noticeably darker cuticles than those

infected with Ac-PG at 24 h post-mortem. The results are representative of three biological replicates. To analyze caspase activity, Sf9 cells were infected with the viruses shown and caspase activity was measured at D) 24 h p.i. and E) 48 h p.i. HA-tagged Hesp018 (Ac-*hesp018-ha*-REP-PG) or untagged Hesp018 (Ac-*hesp018* -REP-PG) were expressed from a version of AcMNPV lacking the caspase inhibitor *p35*. As a comparison, cells were infected with the parental *p35* mutant virus (Ac-p35KO-PG) or a repair virus in which *p35* was reinserted in the *p35* mutant bacmid (Ac-p35-REP). The results shown are combined from four biological replicates. Standard errors and statistical differences obtained by unpaired two-tailed Student's t test are indicated (*p* values are indicated as described in the Fig. 2 legend).

Some serpins, including the poxvirus serpin crmA, are known to be able to inhibit caspases, even though caspases are cysteine proteinases (Tewari *et al.*, 1995). To test whether Hesp018 expression could inhibit caspases during AcMNPV infection, we expressed Hesp018 in a version of AcMNPV lacking P35 (Huang *et al.*, 2011). Infection of Sf9 cells with *p35* mutant AcMNPV resulted in high levels of effector caspase activity, as previously shown (Bertin *et al.*, 1996; Huang *et al.*, 2013). Expression of Hesp018 from the *p35* mutant virus resulted in a significant reduction in caspase activity, although the level of caspase activity was not reduced as much as when *p35* was re-inserted into the *p35* deletion virus (Fig. 5D-E). Nevertheless, these results indicate that Hesp018 may have some inhibitory activity against caspases, although this will require additional confirmation.

3.5. Hesp018 expression increases AcMNPV virulence in *T. ni*

To examine the effect of serpin expression during *in vivo* infection, we performed neonate oral infection assays using the Hesp018-expressing virus and control virus in two different species (*Spodoptera frugiperda* and *Trichoplusia ni*). While both species are susceptible to AcMNPV infection, *T. ni* is more sensitive than *S. frugiperda*, thus providing a valuable comparison. We found no significant difference in LT_{50} (time necessary for 50% lethality) for either *T. ni* or *S. frugiperda* neonates using an OB concentration sufficient to kill 95% of the insects (Table 2). Although the slopes (indicating steepness of the lethality curves) showed statistical differences in both species, the slope for the Hesp018-expressing virus was greater than control virus in *T. ni*, but lower than control in *S. frugiperda*, indicating no clear trend.

However, when the LC_{50} (concentration of occlusion bodies required for 50% lethality) values of these viruses were compared, the LC_{50} of the Hesp018-expressing virus was 4-fold less than the control virus in *T.ni* larvae (Table 3). No significant difference was observed in *S. frugiperda* (Table 3). Thus, expression of Hesp018 in AcMNPV resulted in an increase in viral virulence in *T. ni* larvae.

Table 2. Time-mortality response of *T. ni*and *S. frugiperda* neonate larvae infected orally with either Ac-PGor Ac-*hesp018*-PG.

Insect	Virus	LT_{50} (h)	95% Fiducial limits		Slope±SE
			Lower	Upper	
T. ni	Control	70.87983	64.99885	76.58491	9.8535 ± 0.9100
	Serpin	75.87458	70.98130	80.71219	11.8885 ± 1.0170
S. frugiperda	Control	88.35941	84.47834	92.24068	18.2834 ± 1.4670
	Serpin	82.29592	78.07245	86.57326	15.7703 ± 1.2303

Table 3. Dose-mortality response of *T. ni*and *S. frugiperda* neonate larvae infected orally with either Ac-PG orAc-*hesp018*-PG.

Insect	Virus	LC_{50} (OB/ml)	95% Fiducial limits		Slope ± SE
			Lower	Upper	
T. ni	Control	6.04×10^3	4.19×10^3	8.41×10^3	1.706 ± 0.264
	Serpin	1.55×10^3	0.73×10^3	2.51×10^3	1.239 ± 0.213
S. frugiperda	Control	1.21×10^6	4.45×10^5	2.18×10^6	1.012 ± 0.184
	Serpin	6.57×10^5	1.51×10^5	1.37×10^6	0.9590.197

4. Discussion

The serpin homolog *hesp018*, previously reported in the genome of the Group II alphabaculovirus HespNPV (Rohrmann *et al.*, 2013), is the first intact serpin gene found in a viral genome outside of the chordopoxviruses. In this report, we investigated the inhibitory activity of Hesp018 on several serine proteinases, and showed that it is a functional serpin. Expression of *hesp018* in AcMNPV resulted in accelerated BV production in Sf9 cells and a 4-fold lower LC_{50} in *T. ni* larvae. These results support the hypothesis that acquisition of a serpin homolog provided an evolutionary advantage to an ancestor of HespNPV, causing it to be retained in this lineage. Although the natural host(s) of HespNPV is not known for certain, it was likely isolated from a *Hemileuca sp.* in the family Saturniidae, which is grouped together with Sphingidae and Bombycidae in the superfamily Bombycoidea (Regier *et al.*, 2008). However, the *hesp018* sequence is relatively divergent from noctuid and bombycoid lepidopteran serpin-4 homologs, indicating that if it was acquired from a bombycoid host, it is either not a recent acquisition, or it has evolved faster than the insect genes (Fig. 1A).

At this point we do not know the target enzyme(s) that are acted upon by Hesp018 during HespNPV infection. Serpins are able to maintain the PO cascade in an inactive state in the absence of immune challenge, and down-regulate the cascade during or after infection (Jiang *et al.*, 2010). One of the closest relatives to hesp018, serpin-4 from *M. sexta*, can inhibit hemolymph proteinases and block PO cascade activation (Tong *et al.*, 2005; Tong & Kanost, 2005). Therefore, we hypothesized that the *hesp018* might play a role in regulating host immunity. Our data do indeed suggest that Hesp018 can suppress activation of the PO cascade, since incubation of recombinant Hesp018 with *M. sexta* hemolymph inhibited PO activity. However, it may very well be that the evolutionarily important function of Hesp018 is to inhibit another cellular proteinase that is not involved in PO activation.

Interestingly, expression of either Hesp018 or *M. sexta* serpin-4B caused increased cathepsin activity in infected Sf9 cells. The mechanism of this increase in activity is unknown, but the simplest explanation is that it could be due to inhibition of a proteinase that normally degrades v-cath, since uninfected Sf9 cells or cells infected with a virus lacking v-cath have very low levels of endogenous cathepsin activity (Fig. 5A). However it is also possible that this increase in activity is due to increased expression or activity of a cellular cathepsin. AcMNPV infection causes late melanization of larvae, usually after the host has died, but viruses lacking v-cath do not cause melanization (Slack *et al.*, 1995). Consistent with this, and also with the increased cathepsin activity observed *in vitro*, *M. sexta* larvae infected with AcMNPV expressing Hesp018 melanized more rapidly than controls. Since melanization is a result of PO activation, this more rapid melanization response may seem incongruent with the ability of recombinant Hesp018 to inhibit the PO response. However, the

melanization of the infected larvae occurred late in infection, after death. Presumably at these late times, overwhelming PO activation occurs that cannot be inhibited by expression of Hesp018. Increased levels or activity of v-cath would be expected to accelerate this late activation.

Like other large DNA viruses, baculoviruses have frequently acquired host genes during their evolution, so from this point of view the horizontal acquisition of a proteinase inhibitor is not particularly surprising (Becker, 2000; Katsuma *et al.*, 2012). It is curious, however, that acquisition of serpin homologs has been so rare during virus evolution. Well over 60 baculovirus genomes have been sequenced, but no other baculoviruses have been found to date that harbor serpin homologs, even though these viruses have co-evolved with their lepidopteran hosts for more than 100 million years (Theze *et al.*, 2011). For that matter, functional serpins have not been found in any other viral genomes outside of the vertebrate-infecting chordopoxviruses. The scarcity of serpin homologs in baculoviruses (as well as other viruses) suggests that serpin expression may confer an advantage only in rare situations. It is even possible that serpin expression could have deleterious effects on viral fitness in many situations. For example, in the case of insect viruses such as baculoviruses, serpin expression could potentially allow increased competition by other microbes if inhibiting PO activation results in humoral immunity becoming compromised. Many chordopoxviruses encode multiple serpin homologs (Haller *et al.*, 2014), and serpin expression contributes to the exquisite abilities of these viruses to manipulate the vertebrate immune response. Despite this, other vertebrate-infecting DNA viruses that also inhibit immune responses, such as herpes viruses and adenoviruses, have not acquired serpin homologs. Interestingly, intraperitoneal delivery of purified Serp-1 protein from myxomavirus to

mice infected with gammaherpesvirus 68 or ebolavirus improved host survival and reduced viral infection (Chen *et al.*, 2013). Although this is an artificial situation, it is consistent with the idea that serpin expression may only be advantageous to viruses in highly specialized situations.

Along these lines, in this work we used three different insect species from two lepidopteran families, the two noctuids *T. ni* and *S. frugiperda* and the sphingid *M. sexta*, to characterize the function of *hesp018*. However, even within the order Lepidoptera, immune responses have been found to vary at the family level. For example, the expression of hemolin, an immunoglobulin-like protein specific to lepidopterans, is stimulated by baculovirus infection in bombycoids (*Antheraea pernyi* and *Hyalophora cecropia* from Saturniidae, and *Bombyx mori* from Bombycidae) (Li *et al.*, 2005) but not in the noctuids *Helicoverpa zea*, *Heliothis virescens*, or *T. ni* (Terenius *et al.*, 2009). Interestingly, knocking down hemolin expression accelerated baculovirus infection in *A. pernyi* (Hirai *et al.*, 2004). Importantly, whereas hemolymph from a noctuid (*H. virescens*) exhibited virucidal activity (Popham *et al.*, 2004; Shelby & Popham, 2006), hemolymph proteins from a saturniid (*Lonomia obliqua*) were actually able to improve baculovirus replication *in vitro* (Sousa *et al.*, 2014). Furthermore, encapsulation and melanization of AcMNPV-infected cells by hemocytes was shown to occur in semi-permissive *H. zea* but not in fully permissive *H. virescens* (Trudeau *et al.*, 2001). Therefore, not just hemolin but possibly other components of the lepidopteran hemolymph could play differing roles in the protection of the host insect in different lepidopteran families.

In conclusion, we have shown that the baculovirus-encoded serpin Hesp018 is an active serpin, and that expression of Hesp018 in the heterologous baculovirus AcMNPV provided a replication advantage *in vitro* and enhanced virulence *in vivo* in one of two noctuid hosts. We can only speculate that perhaps the natural host of HespNPV is a lepidopteran species that is adept at mounting a humoral immune response that inhibits baculovirus infection, or some other response that Hesp018 can inhibit, and thus retention of *hesp018* confers a unique advantage to this virus. It would be interesting to study the function of *hesp018* in its natural context, if the natural host could be identified. It will also be interesting to discover, as new viral genomes continue to be sequenced, whether other serpins have been acquired during virus evolution, and their roles in viral replication.

5. Methods

5.1. Cells, virus, and insects

S. frugiperda (fall armyworm) Sf9 and *T. ni* (cabbage looper) TN-368 cells were cultured at 27°C in TC-100 medium (Invitrogen) supplemented with 10% fetal bovine serum, penicillin G (60 µg/ml), streptomycin sulfate (200 µg/ml), and amphotericin B (0.5 µg/ml).Viruses were titered by $TCID_{50}$ assay (O'Reilly *et al.*, 1992) in Sf9 or TN-368 cells (for *p35*-deleted viruses). *M. sexta* eggs were obtained from Michael Kanost, Kansas State University. Insect larvae were reared as described previously (Dunn & Drake, 1983). *T. ni* and *S. frugiperda* eggs were purchased from Benzon Research (Carlisle, PA). After hatching, larvae were reared on artificial diet in a 27 °C chamber with a 12 h/12 h light/dark cycle.

5.2. Gene amplification and construction of shuttle vectors and recombinant viruses

The *hesp018* gene (with or without HA tag) or the *M. sexta serpin-4B* gene were separately amplified using different sets of primers (0.4 µM): hesp018/SacI F (GAG CTC ATG AAC ATG GTC GTC GGA TCA TCG TTA ATA G) and hesp018/NotI-HA R (GCG GCC GCT TAA GCG TAA TCC GGG ACG TCG TAG GGA TAA TTT GAA ATG AAA TCC ATT TTC GTC G) or hesp018/NotI R GCG GCC GCT TAA TTT GAA ATG AAA TCC ATT TTC GTC G); Serpin 4 A/B F (GGA TCC GAG CTC ATG AAG TGT GTG TTA GTG ATT GTA TTA TG) and Serpin 4 A/B R (GGA TCC GCG GCC GCT TAG TAA AGA AAA GGT TGT TTG TAT ATT CC). The amplified fragments were digested with SacI/NotI (New England Biolabs) and cloned into the shuttle plasmid pFB-PG-H-pA (a modified pFB-PG (Wu *et al.*, 2006) containing a SV40-polyA signal and the *Drosophila melanogaster* hsp70 promoter to drive heterologous gene expression). The modified plasmids containing the *serpin* genes were transformed into DH10-Bac cells (Invitrogen, Carlsbad, CA, USA) and recombinant bacmids were selected and confirmed by PCR. Moreover, the plasmids were transposed into both a *cathepsin/chitinase*-deleted (Kaba *et al.*, 2004) and a *p35*-deleted bacmid (Huang *et al.*, 2011). 1 µg of each recombinant bacmid was transfected into Sf9 cells (10^6) using Lipofectin. For the *p35*-deleted bacmid, the transfection was performed using TN-368 cells. The supernatants containing the recombinant viruses were collected at seven days post-transfection, amplified in Sf9 or TN-368 cells, and titered.

5.3. Phylogenetic analysis

A MAFFT alignment (Katoh *et al.*, 2002) was performed with the amino acid sequences of Hesp018 (NC_021923) and serpins from 19 arthropod species (*M. sexta* serpin-4B (AY566163.1); *M. sexta* serpin-4A (AY566162.1); *Bombyx mori* serpin-4 (NM_001043625.1); *Danaus plexippus* (AGBW01006202.1); *Plutella xylostella* serpin-4 (KC686693.1); *Chilo suppressalis* (AFQ01142.1); *B. mori* serpin-5 (NP_001037205.1); *M. sexta* serpin-5 (AY566166.1); *D. plexippus* serpin-5 (EHJ70286.1); *P. xylostella* serpin-5 (AGK24648.1); *B. mandarina* serpin-7 (NP_001139701.1); *M. sexta* serpin-7 (HQ149330.2); *Culex quinquefasciatus* (XM_001863294.1); *Aedes aegypti* (XM_001661855.1); *Anopheles gambiae* (XM_312891.3); *Musca domestica* (XM_005183439.1); *Tribolium castaneum* (XM_008195262.1); *Locusta migratoria* (AGC84400.1); *Caligus rogercresseyi* (BT076733.1). A maximum likelihood tree was inferred using RaxML method with 100 repetitions of a non-parametric bootstrap (Stamatakis *et al.*, 2008) and JTT model selected by Prottest 2.4 (Abascal *et al.*, 2005). The crustacean *C. rogercresseyi* serpin sequence was used to root the tree.

5.4. Serpin expression and purification

The *hesp018* gene was amplified using the primers hesp018 F (GGATCCCATTTAGACCATTTTTCATTAAA) and hesp018 R (AAGCTTAATTTGAAATGAAATCCATTTTC) and HespNPV genomic DNA as template. The generated fragment was cloned into pET19b (Novagen) in BamHI/HindIII restriction sites. The plasmid pET19b-*hesp018* was transformed into *E.*

coli strain BL21(DE3)/pLysS. Recombinant N-terminally His-tagged protein was isolated using Talon resin (Clontech) as previously described (Wu & Passarelli, 2010). The purified protein was dialyzed against phosphate buffer (10 mM NaH2PO4, pH 6.2) and the concentration obtained by BCA assay (Pierce).

5.5. Hemolymph samples and proPO activity inhibition

Fifth-instar, day 3 larvae were chilled on ice for at least 20 min. Hemolymph was collected by clipping the dorsal horn with scissors. Hemocytes were removed by centrifugation at 10,000 × *g* for 10 min at 4 °C. Plasma samples were stored at −80 °C. For proPO activity analysis, 100 ng recombinant protein was incubated with 4 µl hemolymph for 10 min at room temperature. Subsequently, 2 µl of *Micrococcus luteus* extract (10 µg/µl in sterile water, Sigma) was added to stimulate PO activity. 300 ng of BSA was used as a negative control. After incubation for 10 min at room temperature, PO activity was measured by absorbance using dopamine as substrate. One unit of PO activity was defined as the amount of enzyme producing an increase in absorbance (A470) of 0.001 per min. Treatments were replicated three times and analyzed by unpaired two-tailed Student's t test and one-way ANOVA.

5.6. *M. sexta* injection

Fifth-instar, day 3 larvae were chilled on ice for at least 20 min. Three caterpillars were each injected with 100 µl of virus (10^7pfu/ml). For this experiment, we used the *cathepsin/chitinase*-deleted virus transposed with shuttle vectors harboring no gene, *M. sexta*-derived *serpin-4B* gene, or HA-tagged *hesp018* gene. After injection, the insects

were individually transferred into 1-oz plastic cups containing food. Photographs of dead caterpillars were taken 24 h post-death.

5.7. Amidase activity

Recombinant Hesp018 protein was incubated with proteinases at different molar ratios as described (An *et al.*, 2012). One unit of amidase activity was defined as the amount of enzyme producing an increase in absorbance (A405) of 0.001 per min. Treatments were replicated three times and analyzed by one-way ANOVA. The following proteinases and their artificial substrates (Sigma-Aldrich, St. Louis, MO, USA) were used: bovine pancreatic α-chymotrypsin (120 ng) and N-succinyl-Ala-Ala-Pro-Phe-p-nitroanilide; human serum plasmin (200 ng) and D-Phe-L-Pro-L-Arg-p-nitroanilide; proteinase K from *Tritirchium album* (40 ng) (Promega, Madison, WI, USA) and N-succinyl-Ala-Ala-Pro-Phe-p-nitroanilide (synthesized in-house); bovine pancreatic trypsin (5 ng) and N-acetyl-Ile-Glu-Ala-Arg-p-nitroanilide (Biochemistry Core Facility, Kansas State University). Enzymes and substrates were kindly provided by Drs. Michael Kanost and Kristin Michel, Kansas State University.

5.8. Secretion analysis

Sf9 cells were infected at MOI of 5 with Ac-PG, Ac-*hesp018*-ha-PG, or Ac-*Ms-serpin-4B*-PG for 1 hour at 27ºC, washed three times with fresh media, and replaced with 1 ml of TC-100 without FBS. At different time points, cells and supernatant were collected and centrifuged at 1,000 x *g* for 5 min at room temperature. The supernatant was transferred to a new tube and the pelleted cells were washed twice with phosphate-

buffered saline (PBS), pH 6.2 and resuspended in PBS. The supernatant and the resuspended cells were incubated with an equal volume of 2x protein loading buffer and heated for 5 min at 100 ºC. Proteins were analyzed by SDS-PAGE (sodium dodecyl sulfate-polyacrylamide gel electrophoresis) followed by immunoblotting using monoclonal anti-HA (Covance) or anti-*M. sexta* Serpin-4B (Tong & Kanost, 2005) kindly provided by Michael Kanost.

5.9. Viral growth curves

Sf9 cells were infected at MOI=0.01. After 1 hr, virus was removed, the cells were washed twice with TC-100, and TC-100 containing 10% FBS was added. Samples were collected at the indicated times and titered by $TCID_{50}$ assay.

5.10. Cathepsin and chitinase activity

Sf9 cells were infected at an MOI=5 with Ac-PG, Ac-*hesp018*-ha-PG, Ac-Ms-serpin-4B-PG, or AcDelCC-PG (Kaba *et al.*, 2004). At 42 h post infection (p.i.), cells and supernatant were collected and centrifuged at 500 x *g* for 5 min at 4º C. The cells were washed twice with PBS. The final pellet was resuspended in 500 µl of PBS. The cells were lysed on ice using a glass homogenizer. The protein concentrations were obtained by BCA assay (Pierce), and 100 µg (cathepsin) or 10 µg (chitinase) of lysate was used for activity assays as previously described (Gopalakrishnan *et al.*, 1995; Slack *et al.*, 1995).

5.11. Caspase activity

Sf9 cells were infected at MOI=5 with *p35*-deleted viruses harboring the *hesp018* gene with or without HA tag or *p35* for 1 h at 27 ºC. At 24 and 48 h p.i., cells were collected and washed twice with PBS and resuspended in 100 µl lysis buffer (20 mM HEPES KOH, pH 7.5, 50 mM KCl, 1.5 mM MgCl2, 1 mM EDTA, 1 mM EGTA, 1 mM DTT, 250 mM sucrose) containing Complete Mini EDTA-free proteinase inhibitor cocktail. 50 µl of cell lysate was incubated with 50 µl of reaction buffer (100 mM HEPES, pH 7.4, 2 mM DTT, 0.1% CHAPS, 1% sucrose) at 37 ºC for 15 min. Caspase substrate Ac-DEVD-AFC (MP Biomedicals) was added at a final concentration of 40 µM and the fluorescence (excitation wave length of 405 nm and emission wave length of 505 nm) was monitored every 10 min for 2 h at 25 ºC using a Victor 1420 Multilabel counter (Perkin-Elmer). The average slope (change in fluorescence versus time) was plotted.

5.12. Bioassays in *T. ni* and *S. frugiperda* neonates

T. ni and *S. frugiperda* neonates (within 24 h after hatching) were transferred to diet contaminated with OBs at different concentrations. OBs from Ac-PG and Ac-*hesp018*-PG were obtained from *per os*-infected *T. ni*, purified (O'Reilly *et al.*, 1992), resuspended in water, and vortexed for 2 h to dissociate clumps. Concentrations of OBs at 1.6×10^2, 8.0×10^3, 4.0×10^4, 2.0×10^5, and 1.0×10^6 OBs/ml for *T. ni* or 8.0×10^5, 4.0×10^6, 2.0×10^7, and 1.0×10^8 OBs/ml for *S. frugiperda* in the diet were used, as previously described (Detvisitsakun *et al.*, 2007). After feeding for 24 h, neonates were transferred into individual plastic cups containing uncontaminated food. Mortality was recorded at different time points by scoring the number of dead insects which had no

response to touch. The LC_{50} and LT_{50} values were determined using probit analysis (SAS Institute, 2004).

6. Acknowledgements

We thank Kristin Michel, Xin Zhang, Michael Kanost, and Daisuke Takahashi (Kansas State University) for reagents, advice, and helpful discussion. We are also grateful to Monique van Oers (Wageningen University) for providing the *chitinase*/*cathepsin* deletion virus. D.M.P.A.-A. was supported by the Brazilian funding agency 'Coordenação de Aperfeiçoamento de Pessoal de Nível Superior' (CAPES). This work was supported in part by the Kansas Agricultural Experiment Station. Contribution no. 15-189-J from the Kansas Agricultural Experiment Station.

7. References

Abascal, F., Zardoya, R. & Posada, D. (2005). ProtTest: selection of best-fit models of protein evolution. *Bioinformatics* **21**, 2104-2105.

An, C., Hiromasa, Y., Zhang, X., Lovell, S., Zolkiewski, M., Tomich, J. M. & Michel, K. (2012). Biochemical characterization of Anopheles gambiae SRPN6, a malaria parasite invasion marker in mosquitoes. *PLoS One* **7**, e48689.

Becker, Y. (2000). Evolution of viruses by acquisition of cellular RNA or DNA nucleotide sequences and genes: an introduction. *Virus Genes* **21**, 7-12.

Bertin, J., Mendrysa, S. M., LaCount, D. J., Gaur, S., Krebs, J. F., Armstrong, R. C., Tomaselli, K. J. & Friesen, P. D. (1996). Apoptotic suppression by baculovirus P35 involves cleavage by and inhibition of a virus-induced CED-3/ICE-like protease. *J Virol* **70**, 6251-6259.

Chen, H., Zheng, D., Abbott, J., Liu, L., Bartee, M. Y., Long, M., Davids, J., Williams, J., Feldmann, H., Strong, J., Grau, K. R., Tibbetts, S., Macaulay, C., McFadden, G., Thoburn, R., Lomas, D. A., Spinale, F. G., Virgin, H. W. & Lucas, A. (2013). Myxomavirus-derived serpin prolongs survival and reduces inflammation and hemorrhage in an unrelated lethal mouse viral infection. *Antimicrob Agents Chemother* **57**, 4114-4127.

Clem, R. J. (2005). The role of apoptosis in defense against baculovirus infection in insects. *Curr Top Microbiol Immunol* **289**, 113-129.

Clem, R. J. & Passarelli, A. L. (2013). Baculoviruses: sophisticated pathogens of insects. *PLoS Pathog* **9**, e1003729.

Dean, P., Potter, U., Richards, E. H., Edwards, J. P., Charnley, A. K. & Reynolds, S. E. (2004). Hyperphagocytic haemocytes in Manduca sexta. *J Insect Physiol* **50**, 1027-1036.

Detvisitsakun, C., Cain, E. L. & Passarelli, A. L. (2007). The Autographa californica M nucleopolyhedrovirus fibroblast growth factor accelerates host mortality. *Virology* **365**, 70-78.

Dunn, P. E. & Drake, D. R. (1983). Fate of bacteria injected into naive and immunized larvae of the tobacco hornworm Manduca sexta. *Journal of Invertebrate Pathology* **41**, 9.

Ferrandon, D., Imler, J. L., Hetru, C. & Hoffmann, J. A. (2007). The Drosophila systemic immune response: sensing and signalling during bacterial and fungal infections. *Nat Rev Immunol* **7**, 862-874.

Gettins, P. G. (2002). Serpin structure, mechanism, and function. *Chem Rev* **102**, 4751-4804.

Gopalakrishnan, B., Muthukrishnan, S. & Kramer, K. J. (1995). Baculovirus-mediated Expression of a *Manduca sexta* Chitinase Gene: Properties of the Recombinant Protein. *Insect Biochem Mol Biol* **25**, 11.

Haller, S. L., Peng, C., McFadden, G. & Rothenburg, S. (2014). Poxviruses and the evolution of host range and virulence. *Infect Genet Evol* **21**, 15-40.

Hawtin, R. E., Zarkowska, T., Arnold, K., Thomas, C. J., Gooday, G. W., King, L. A., Kuzio, J. A. & Possee, R. D. (1997). Liquefaction of Autographa californica nucleopolyhedrovirus-infected insects is dependent on the integrity of virus-encoded chitinase and cathepsin genes. *Virology* **238**, 243-253.

Herniou, E. A., Arif, B. M., Becnel, J. J., Blissard, G. W., Bonning, B., Harrison, R., Jehle, J. A., Theilmann, D. A. & Vlak, J. M. (2012). Baculoviridae. In *Virus taxonomy Ninth Report of the International Committee on Taxonomy of Viruses* pp. 163-173. Edited by A. M. Q. King, M. J. Adams, E. B. Carstens & E. J. Lefkowitz. San Diego: Elsevier-Academic Press.

Hirai, M., Terenius, O., Li, W. & Faye, I. (2004). Baculovirus and dsRNA induce Hemolin, but no antibacterial activity, in Antheraea pernyi. *Insect Mol Biol* **13**, 399-405.

Huang, N., Civciristov, S., Hawkins, C. J. & Clem, R. J. (2013). SfDronc, an initiator caspase involved in apoptosis in the fall armyworm Spodoptera frugiperda. *Insect Biochem Mol Biol* **43**, 444-454.

Huang, N., Wu, W., Yang, K., Passarelli, A. L., Rohrmann, G. F. & Clem, R. J. (2011). Baculovirus infection induces a DNA damage response that is required for efficient viral replication. *J Virol* **85**, 12547-12556.

Huntington, J. A. (2011). Serpin structure, function and dysfunction. *J Thromb Haemost* **9 Suppl 1**, 26-34.

Ikeda, M., Yamada, H., Ito, H. & Kobayashi, M. (2011). Baculovirus IAP1 induces caspase-dependent apoptosis in insect cells. *J Gen Virol* **92**, 2654-2663.

Jakubowska, A. K., Vogel, H. & Herrero, S. (2013). Increase in gut microbiota after immune suppression in baculovirus-infected larvae. *PLoS Pathog* **9**, e1003379.

Jayachandran, B., Hussain, M. & Asgari, S. (2012). RNA interference as a cellular defense mechanism against the DNA virus baculovirus. *J Virol* **86**, 13729-13734.

Jiang, H., Vilcinskas, A. & Kanost, M. R. (2010). Immunity in lepidopteran insects. *Adv Exp Med Biol* **708**, 181-204.

Kaba, S. A., Salcedo, A. M., Wafula, P. O., Vlak, J. M. & van Oers, M. M. (2004). Development of a chitinase and v-cathepsin negative bacmid for improved integrity of secreted recombinant proteins. *J Virol Methods* **122**, 113-118.

Kamita, S. G., Nagasaka, K., Chua, J. W., Shimada, T., Mita, K., Kobayashi, M., Maeda, S. & Hammock, B. D. (2005). A baculovirus-encoded protein tyrosine phosphatase gene induces enhanced locomotory activity in a lepidopteran host. *Proc Natl Acad Sci U S A* **102**, 2584-2589.

Katoh, K., Misawa, K., Kuma, K. & Miyata, T. (2002). MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res* **30**, 3059-3066.

Katsuma, S., Koyano, Y., Kang, W., Kokusho, R., Kamita, S. G. & Shimada, T. (2012). The baculovirus uses a captured host phosphatase to induce enhanced locomotory activity in host caterpillars. *PLoS Pathog* **8**, e1002644.

Li, W., Terenius, O., Hirai, M., Nilsson, A. S. & Faye, I. (2005). Cloning, expression and phylogenetic analysis of Hemolin, from the Chinese oak silkmoth, Antheraea pernyi. *Dev Comp Immunol* **29**, 853-864.

McNeil, J., Cox-Foster, D., Slavicek, J. & Hoover, K. (2010). Contributions of immune responses to developmental resistance in Lymantria dispar challenged with baculovirus. *J Insect Physiol* **56**, 1167-1177.

Nappi, A. J. & Christensen, B. M. (2005). Melanogenesis and associated cytotoxic reactions: applications to insect innate immunity. *Insect Biochem Mol Biol* **35**, 443-459.

O'Reilly, D., Miller, L. K. & Luckow, V. A. (1992). *Baculovirus Expression Vectors: a laboratory manual*. New York: Freeman and Company.

O'Reilly, D. R. & Miller, L. K. (1989). A baculovirus blocks insect molting by producing ecdysteroid UDP-glucosyl transferase. *Science* **245**, 1110-1112.

Ooi, B. G. & Miller, L. K. (1988). Regulation of host RNA levels during baculovirus infection. *Virology* **166**, 515-523.

Popham, H. J., Shelby, K. S., Brandt, S. L. & Coudron, T. A. (2004). Potent virucidal activity in larval Heliothis virescens plasma against Helicoverpa zea single capsid nucleopolyhedrovirus. *J Gen Virol* **85**, 2255-2261.

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

Regier, J. C., Cook, C. P., Mitter, C. & Hussey., A. (2008). A phylogenetic study of the 'bombycoid complex' (Lepidoptera) using five protein-coding nuclear genes, with comments on the problem of macrolepidoteran phylogeny. *Systematic Entomology* **33**, 15.

Rohrmann, G. F. (2013a). *Baculovirus Molecular Biology*. Bethesda (MD): National Center for Biotechnology Information (US). Available from: http://www.ncbi.nlm.nih.gov/books/NBK49500/.

Rohrmann, G. F. (2013b). Characterization of baculoviruses from the Martignoni collection. *J Invertebr Pathol* **114**, 61-64.

Rohrmann, G. F., Erlandson, M. A. & Theilmann, D. A. (2013). The genome of a baculovirus isolated from Hemileuca sp. encodes a serpin ortholog. *Virus Genes* **47**, 357-364.

Shelby, K. S. & Popham, H. J. (2006). Plasma phenoloxidase of the larval tobacco budworm, Heliothis virescens, is virucidal. *J Insect Sci* **6**, 1-12.

Silverman, G. A., Bird, P. I., Carrell, R. W., Church, F. C., Coughlin, P. B., Gettins, P. G., Irving, J. A., Lomas, D. A., Luke, C. J., Moyer, R. W., Pemberton, P. A., Remold-O'Donnell, E., Salvesen, G. S., Travis, J. & Whisstock, J. C. (2001). The serpins are an expanding superfamily of structurally similar but functionally diverse proteins. Evolution, mechanism of inhibition, novel functions, and a revised nomenclature. *J Biol Chem* **276**, 33293-33296.

Slack, J. & Arif, B. M. (2007). The baculoviruses occlusion-derived virus: virion structure and function. *Adv Virus Res* **69**, 99-165.

Slack, J. M., Kuzio, J. & Faulkner, P. (1995). Characterization of v-cath, a cathepsin L-like proteinase expressed by the baculovirus Autographa californica multiple nuclear polyhedrosis virus. *J Gen Virol* **76 (Pt 5)**, 1091-1098.

Sousa, A. P., Moraes, R. H. & Mendonca, R. Z. (2014). Improved replication of the baculovirus Anticarsia gemmatalis nucleopolyhedrovirus (AgMNPV) in vitro using proteins from Lonomia obliqua hemolymph. *Cytotechnology*, Epub ahead of print.

Stamatakis, A., Hoover, P. & Rougemont, J. (2008). A rapid bootstrap algorithm for the RAxML Web servers. *Syst Biol* **57**, 758-771.

Terenius, O., Popham, H. J. & Shelby, K. S. (2009). Bacterial, but not baculoviral infections stimulate Hemolin expression in noctuid moths. *Dev Comp Immunol* **33**, 1176-1185.

Tewari, M., Telford, W. G., Miller, R. A. & Dixit, V. M. (1995). CrmA, a poxvirus-encoded serpin, inhibits cytotoxic T-lymphocyte-mediated apoptosis. *J Biol Chem* **270**, 22705-22708.

Theze, J., Bezier, A., Periquet, G., Drezen, J. M. & Herniou, E. A. (2011). Paleozoic origin of insect large dsDNA viruses. *Proc Natl Acad Sci U S A* **108**, 15931-15935.

Thiem, S. M. (2009). Baculovirus genes affecting host function. *In Vitro Cell Dev Biol Anim* **45**, 111-126.

Tong, Y., Jiang, H. & Kanost, M. R. (2005). Identification of plasma proteases inhibited by Manduca sexta serpin-4 and serpin-5 and their association with components of the prophenol oxidase activation pathway. *J Biol Chem* **280**, 14932-14942.

Tong, Y. & Kanost, M. R. (2005). Manduca sexta serpin-4 and serpin-5 inhibit the prophenol oxidase activation pathway: cDNA cloning, protein expression, and characterization. *J Biol Chem* **280**, 14923-14931.

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

Washburn, J. O., Haas-Stapleton, E. J., Tan, F. F., Beckage, N. E. & Volkman, L. E. (2000). Co-infection of Manduca sexta larvae with polydnavirus from Cotesia congregata increases susceptibility to fatal infection by Autographa californica M Nucleopolyhedrovirus. *J Insect Physiol* **46**, 179-190.

Wu, W., Lin, T., Pan, L., Yu, M., Li, Z., Pang, Y. & Yang, K. (2006). Autographa californica multiple nucleopolyhedrovirus nucleocapsid assembly is interrupted upon deletion of the 38K gene. *J Virol* **80**, 11475-11485.

Wu, W. & Passarelli, A. L. (2010). Autographa californica multiple nucleopolyhedrovirus Ac92 (ORF92, P33) is required for budded virus production and multiply enveloped occlusion-derived virus formation. *J Virol* **84**, 12351-12361.

Xu, J. & Cherry, S. (2014). Viruses and antiviral immunity in Drosophila. *Dev Comp Immunol* **42**, 67-84.

Yu, X. Q. & Kanost, M. R. (2004). Immulectin-2, a pattern recognition receptor that stimulates hemocyte encapsulation and melanization in the tobacco hornworm, Manduca sexta. *Dev Comp Immunol* **28**, 891-900.

Zhao, P., Li, J., Wang, Y. & Jiang, H. (2007). Broad-spectrum antimicrobial activity of the reactive compounds generated in vitro by Manduca sexta phenoloxidase. *Insect Biochem Mol Biol* **37**, 952-959.

Zhao, P., Lu, Z., Strand, M. R. & Jiang, H. (2011). Antiviral, anti-parasitic, and cytotoxic effects of 5,6-dihydroxyindole (DHI), a reactive compound generated by phenoloxidase during insect immune response. *Insect Biochem Mol Biol* **41**, 645-652.

Capítulo 6. A betabaculovirus encoding a *gp64* homolog

1. Abstract

Background. A betabaculovirus (DisaGV) was isolated from *Diatraea saccharalis* (Lepidoptera: Crambidae), one of the most important insect pest of the sugarcane and other monocot cultures in Brazil. **Results**. The complete genome sequence of DisaGV was determined using the 454-pyrosequencing method. The genome was 98,404 bp long, which makes it the smallest lepidopteran-infecting baculovirus sequenced to date. It had a G+C content of 29.7% encoding 125 putative open reading frames (ORF). All the 37 baculovirus core genes and a set of 19 betabaculovirus-specific genes were found. A group of 13 putative genes was not found in any other baculovirus genome sequenced so far. A phylogenetic analysis indicated that DisaGV is a member of *Betabaculovirus* genus and that it is a sister group to a cluster formed by ChocGV, ErelGV, PiraGV isolates, ClanGV, CaLGV, CypoGV, CrleGV AdorGV, PhopGV and EpapGV. Surprisingly, we found in the DisaGV genome a G protein-coupled receptor related to lepidopteran and other insect virus genes and a *gp64* homolog which is likely a product of horizontal gene transfer from Group 1 alphabaculoviruses. **Conclusion**. DisaGV is a novel species into the genus *Betabaculovirus*. It is closely related to CypoGV-related species and presents the smallest genome in size so far. Remarkably, we found a homolog of *gp64* which used to be present solely in group 1 alphabaculovirus genomes.

Keywords: Baculovirus genome, *Diatraea saccharalis* betabaculovirus, gp64, GPCR, evolution.

Este capítulo ainda não foi publicado. **A betabaculovirus encoding a *gp64* homolog.** Daniel M. P. Ardisson-Araújo[§], Bruna T. Pereira[§], Fernando L. Melo, Bergmann M. Ribeiro, Sônia N. Báo, Paolo M. de A. Zanotto, Flávio Moscardi, Elliot W. Kitajima, Daniel R. Soza-Gomes, José L. C. Wolff.

2. Background

Brazil is the largest sugarcane (*Saccharum officinarum*, L.) and bioethanol producer in the world (Soccol *et al.*, 2010; Zanin *et al.*, 2000). Nowadays, sugarcane is grown on an area over 8 million hectares for both sugar and alcohol production (Soccol *et al.*, 2010). As with other cultures cultivated over large areas, pest control is of crucial importance. The sugarcane borer *Diatraea saccharalis* Fabr. (Lepidoptera: Crambidae) is present in all sugarcane-producing regions of the country, and is considered the major sugarcane pest, especially in the southeast region (Dinardo-Miranda, 2008). Biological control based on the release of the parasitoid *Cotesia flavipes* (Cameron) (Hymenoptera: Braconidae) has been used with success in the control of the sugarcane borer (Mahmoud *et al.*, 2011; Rossi *et al.*, 2014). However, other complementary and compatible methods, such as the application of baculoviruses, would be highly desirable.

Baculoviruses are a large group of insect-specific viruses with circular double-stranded DNA, whose hallmark is the presence of occlusion bodies (OBs)(Rohrmann, 2013). The family *Baculoviridae* comprises four genera: two of them, *Alphabaculovirus* and *Betabaculovirus*, infect insects of the order Lepidoptera; the other two *Gammabaculovirus* and *Deltabaculovirus*, that infect insects of the orders Hymenoptera

and Diptera (Jehle *et al.*, 2006b) respectively. To date more than 100 baculovirus genomes were completely sequenced, and 19 of them are betabaculoviruses.

The Anticarsia gemmatalis multiple nucleopolyhedrovirus (AgMNPV) has been used in Brazil in one of the largest biocontrol programs in the world to control an insect pest (Moscardi, 1999). Other successful programs with baculoviruses have been reported elsewhere in the world (Rohrmann, 2013). The success of the AgMNPV program is due to a combination of factors, such as: high virulence, dead larvae can be collected directly from the field to be used as inoculum, efficient application technology, etc. Nevertheless, development is needed on pest species that are not so easily exposed to the virus, as in the case of borers. Large-scale DNA sequencing provides information on complete viral genomes allowing for "omic" approaches that will eventually facilitate the development of application strategies. Since Brazil is a very diverse country, several baculoviruses have been found and their genomes sequenced (Ardisson-Araujo *et al.*, 2014a; Ardisson-Araujo *et al.*, 2014b; Ardisson-Araujo *et al.*, 2015; Craveiro *et al.*, 2015; Oliveira *et al.*, 2006; Wolff *et al.*, 2008). With this prospect in mind, we have sequenced and analyzed the genome of *Diatraea saccharalis granulovirus* (DisaGV), the first betabaculovirus isolated from a member of the family Crambidae. The presence of a *gp64* homolog was a unique and remarkable finding among betabaculoviruses.

3. Results and Discussion

3.1. Viral infection confirmation

Subjects of the species *Diatraea saccharalis* with virus infection symptoms were found in sugarcane fields in the Southern Brazil. We performed the structural characterization

of the putative virus and a granulovirus infection was confirmed by transmission electron microscopy of OBs extracted from larvae cadavers. Each elliptical granule had a single rod-shaped virion surrounded by a robust protein matrix coat (Figure 1), indicating the typical morphology of GVs (Rohrmann, 2013). Since the protein matrix is formed by granulin produced in large amounts during late infection and because it is highly conserved among lepidopteran-infective baculovirus, we amplified and sequenced the *granulin* gene in order to obtain an initial confirmation to the viral type (data not shown). The 747 bp length of the DisaGV *granulin* had high amino acid identity with orthologs from the genus *Betabaculovirus* (data not shown).

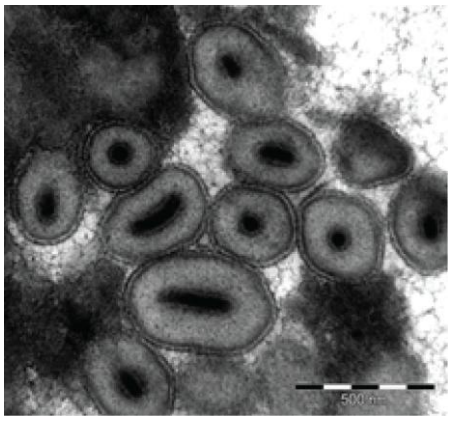

Figure 1. Ultrastructural analysis of *Diatraea saccharalis granulovirus* (DisaGV). Transmission electron micrograph reveals granular occlusion bodies containing singly embedded rod-shaped nucleocapsid (red arrow) (scale bars = 0.5 µm).

3.2. DisaGV genome and phylogeny

The complete genome of DisaGV (Genbank accession number: KP296186) was 98,407 bp in length (mean coverage of 36 x), which makes the DisaGV the smallest

betabaculovirus sequenced to date, followed by AdorGV (99,657 bp) (Wormleaton *et al.*, 2003) and PlxyGV (100,999 bp) (Hashimoto *et al.*, 2000). The G+C content was 29.7 % a typical low value found among GVs and potentially encoded 125 ORFs with at least 50 predicted amino acid residues (Table S1 and Figure 2). The current baculovirus species demarcation criterion is based on pairwise nucleotide distances estimated using the Kimura 2-parameter model of nucleotide substitution for three genes, *granulin*, *lef-8*, and *lef-9* (Jehle *et al.*, 2006b). The pairwise distances of the viral sequences of DisaGV to other betabaculoviruses for both single loci and concatenated alignment are well in excess of 0.05 substitutions/site fulfilling all the criteria for a novel species (data not shown). In order to investigate the phylogenetic relationship of DisaGV to other baculoviruses, we carried out a maximum likelihood phylogenetic analysis based on the alignment of the 37 baculovirus core proteins from all baculovirus genomes publicly available using solely the unique species (Table S2). As suggested by both OB ultrastructural analysis and *granulin* gene sequencing (data not shown), we found DisaGV as sister taxa of the cluster formed by ChocGV, ErelGV, PiraGV isolates, ClanGV, CaLGV, CypoGV and CrleGV (Figure 3A).

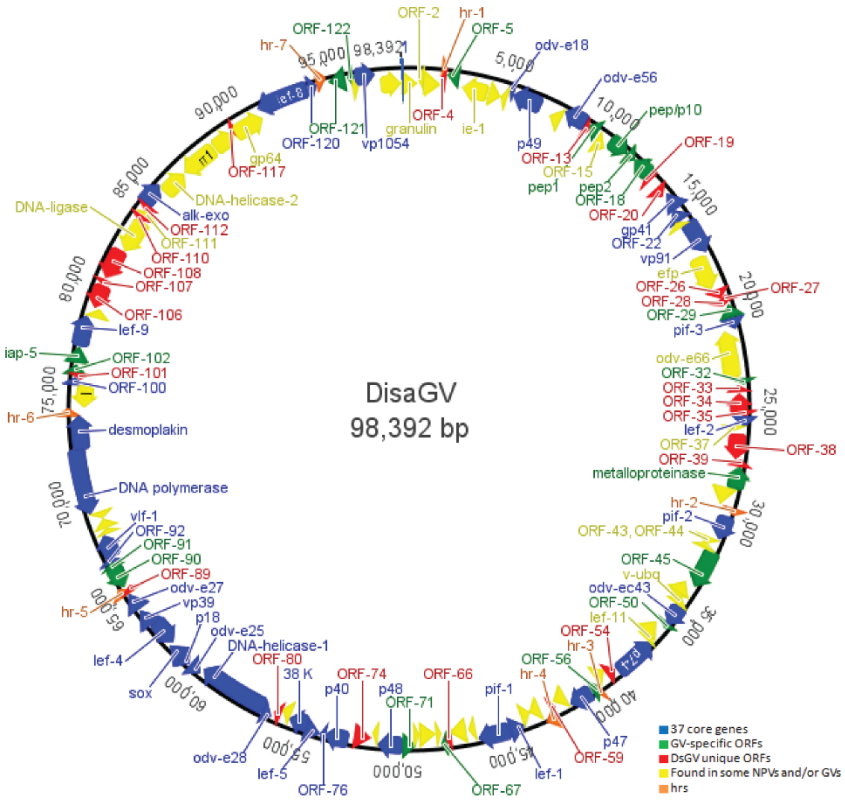

Figure 2. Circular genome map of DisaGV with all genes identified on the 98,392 bp long. Arrows show the transcripcional orientation and relative size of each ORF. Those are colored according to presence into baculoviral genera: in blue the 37 core genes, in green only betabaculovirus-specific genes, in red the DisaGV unique genes, in yellow genes found in some subjects of both alpha and betabaculovirus, and homologous regions (hrs) in orange.

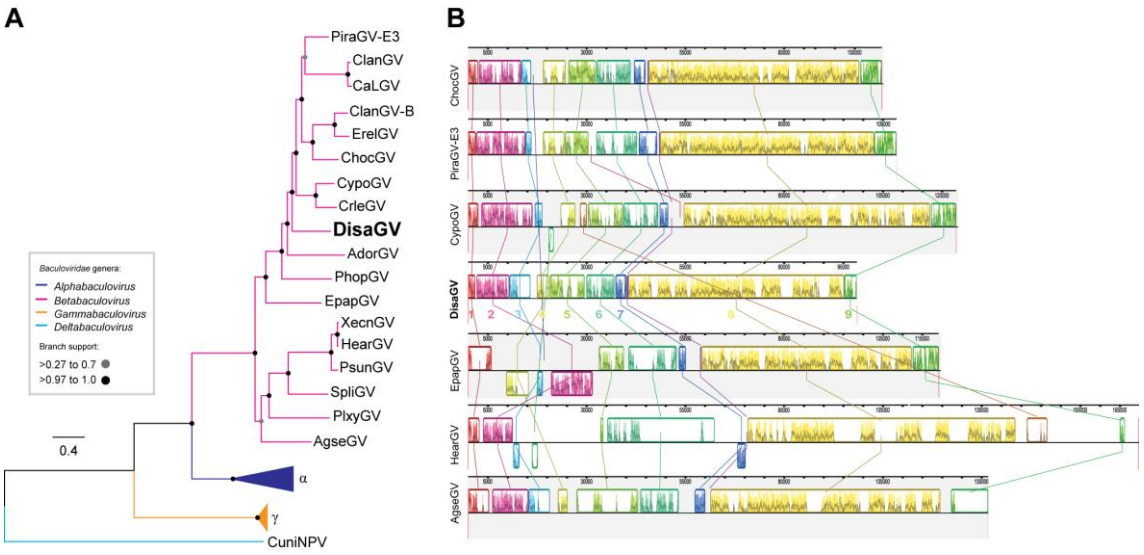

Figure 3. Maximum-likelihood tree for *Betabaculovirus* and genome comparison. (A) The phylogeny was based on the concatenated amino acid sequences of the 37 core proteins identified in all baculovirus genome completely sequenced so far (Table S2). We collapsed gammabaculoviruses (orange, γ) and alphaphabaculoviruses (dark blue, α). The CuniNPV was used as root (light blue). DisaGV (boldface) is a betabaculovirus and sister species of the cluster formed by CypoGV-related species. (B) Genome comparison of the DisaGV genome against some related species including AgseGV, ChocGV, CypoGV, EpapGV, and ErelGV. Locally collinear blocks (LCB) are numbered in the DisaGV genome from 1 to 9. Same colors depict same LCBs across the genomes. Rearrangement can be seen among the species.

Moreover, we performed a genomic comparison among some selected betabaculovirus genomes. We found nine Locally Collinear Blocks (LCB), composed of genomic segments that appear to have the same relative position of their shared genes (Figure 3B). Interestingly, LCB5 (from bp 20013 to 37032), LCB7 (from bp 40326 to 76348) and LCB8 (from bp 76601 to 87652) had an unexpected gene content composition.

LCB5 lacked baculovirus core genes (χ^2=2.46, $p < 0.05$, df =3), while LCB7 had a higher than expected number (χ^2=3.84, $p < 0.05$, df =3) and LCB8 had a higher than expected number of DisaGV-unique genes, a lower than expected number of baculovirus core genes and less than expected GV-specific genes (χ^2= 5.12, $p < 0.01$, df =3).

DisaGV had a *dna-ligase* (*disa107*) and two *helicase* genes (*helicase-1*, *disa081* and *helicase-2*, *disa111*) probably involved in replication, repair, and recombination of DNA (Kuzio *et al.*, 1999). We also identified a *deoxyuridine triphosphatase* (*dut*) gene (*disa073*) and the *ribonucleotide reductase* subunits *rr1* (*disa112*) and *rr2a* (*disa113*), involved in nucleotide metabolism. The role of those genes during baculovirus infection is not clear. It was noteworthy the absence of several genes for early transcription factors, such as the *ie-0*, *ie-2*, and *pe38*. There were also no similar sequences to the *baculovirus repeated ORFs* (*bro* genes), to the *ecdysteroid UDP-glucosyltransferase* (*egt*), to the apoptosis inhibitor *p35*, and also to the *cathepsin* and *chitinase* genes. We observed that the *egt* gene were absent only in the genomes of four other GVs, HearGV, PsunGV, SpliGV-K1 and XecnGV, that form a distinct phylogenic cluster. On the other hand, the *p35* gene was found only in the genomes of ChocGV, CaLGV, ClanGV (Data not shown). The absence of the *cathepsin* and *chitinase* genes may be compensated by the presence of the putative gene for *matrix metalloproteinase* (a *stromelysin-1-like* gene, *disa040*). Whereas the loss of the *cathepsin* and *chitinase* genes is a common event among the betabaculoviruses (Ardisson-Araujo *et al.*, 2014a), the *matrix metalloproteinase* gene is present in all betabaculoviruses sequenced to date (Ishimwe *et al.*, 2015a). The expression a functional CypoGV-encoded metalloproteinase into the

AcMNPV genome enhanced the virus virulence, promoted larval melanization, and could partially substitute the viral cathepsin (Ishimwe *et al.*, 2015b).

3.3. DisaGV unique genes

Homologs to 25 DisaGV ORFs were not found in the genome of other baculoviruses. Taking into account the 450 bp region upstream of each unique ORF, three of them presented no previously characterized promoter motifs, 12 contained exclusively early promoter motifs (TATAW, TATAWAW, TATAWTW with W = A or T), and ten had both early and late (A/TTAAG) motifs (Table S). Two unique ORFs, *disa034* and *disa039* showed significant BlastP hits to other dsDNA virus sequences publicly available. *disa034* encoded a putative 310 aa protein that showed 26% amino acid identity (e-value = 1e-06) to a 247 aa length protein of a phycodnavirus (*Feldmannia irregularis* virus a, AAR26869) (Figure 4A). Moreover, *disa039* coded for a hypothetical protein related to insect-infecting dsDNA viruses including *Wiseana iridescent virus* (WIV) (YP_004732905, 131 aa) and *Amsacta moorei entomopoxvirus* 'L' (NP_064857, 158 aa) (Figure 4B). Phycodnaviruses are eukaryotic algae viruses and seem to share a common ancestor with other insect dsDNA viruses, including iridoviruses and entomopoxviruses, which share baculovirus genes as well (Yamada, 2011). Several baculovirus genes were found into the genome of those viruses, suggesting the occurrence of lateral gene transfer during co-infection in the same insect host, as probably expected to *disa034* (Iyer *et al.*, 2006) and *disa039*.

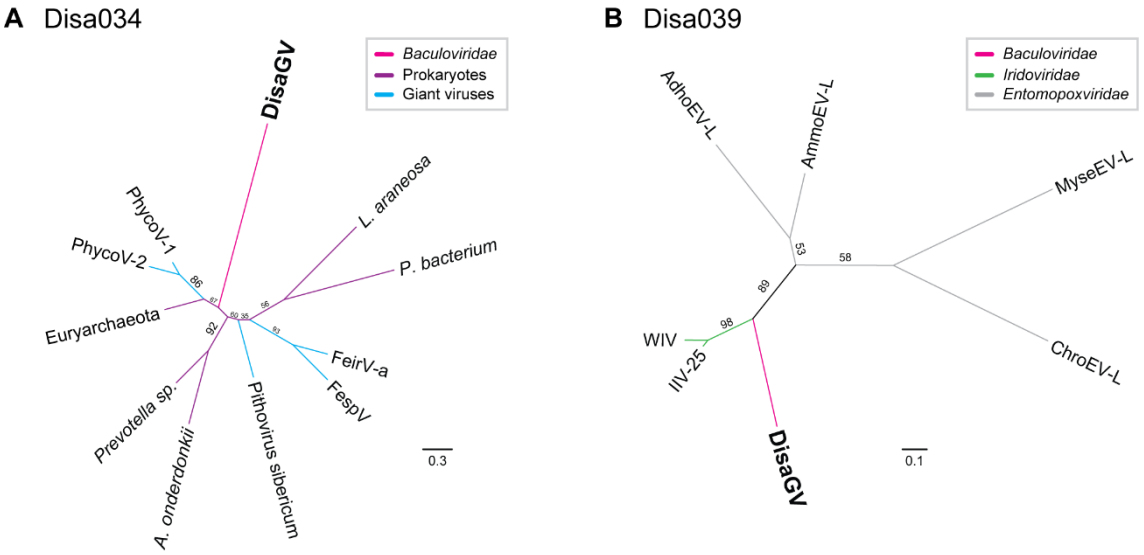

Figure 4. Maximum likelihood phylogenetic trees of both Disa034 (A) and Disa039 (B) based on their predicted amino acid sequence. We used the for RaxML method under the LG+I+G model for Disa034 and WAG+I+F for Disa039 with a nonparametric bootstrap to support the branches. Organisms: (A) Organic Lake phycodnaviruses (PhycoV-1 and PhycoV-2), *Feldmannia species virus* (FespV), *Feldmannia irregularis virus a* (FeirV-a) and Prokaryotes. (B) *Wiseana iridescent virus* (WIV), Invertebrate iridovirus 25 (IIV-25), *Amsacta moorei entomopoxvirus* 'L' (AmmoEV-L), *Adoxophyes honmai entomopoxvirus* 'L' (AdhoEV-L), *Mythimna separata entomopoxvirus* 'L' (MyseEV-L) and *Choristoneura rosaceana entomopoxvirus* 'L' (ChroEV-L).

3.4. G protein-coupled receptor (GPCR)

We also found another unique gene (*disa038*) related to a putative class B secretin-like G-protein coupled receptor (GPCR) of lepidopteran and an entomopoxvirus (Figure 5A). GPCRs are cell membrane-associated GTPases that transmits signals from the environment to the cell inside or between cells allowing them to react to a

corresponding variety of extracellular stimuli that can be mediated by different peptides, lipids, proteins, nucleotides, nucleosides, organic odorants and photons (Kochman, 2014). This type of receptors has been described in many animal species despite of not being quite common in virus genomes We found a predicted signal peptide and seven trans-membrane domains in Disa038 (Figure 5B) making it a member of the Secretin family (Krishnan *et al.*, 2012). Three subfamilies are recognized for this family and one of them, the B2 contains receptors with long extracellular N-termini as observed for both the predicted Disa038 and the other related proteins. It is not clear the role displayed by this gene into DisaGV infection context. Otherwise, the human herpesvirus virus, another dsDNA virus, utilizes virally encoded GPCR to hijack cellular signaling networks for their own benefit suggesting a likely similar pathway during DisaGV infection in the host insect (Nijmeijer *et al.*, 2010).

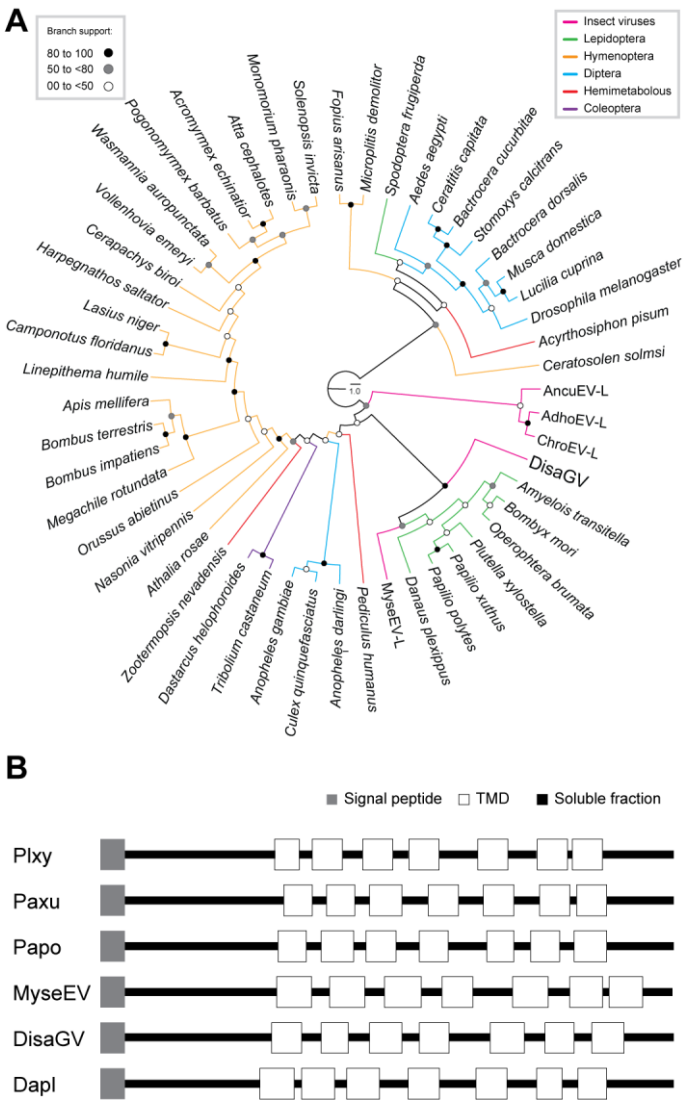

Figure 5. *In silico* analyses of Disa038, a betabaculovirus-encoded G protein-coupled receptor gene. (A) Phylogenetic analysis of selected arthropod GPCRs. Disa038 sequence clustered with lepidopteran and an entomopoxvirus proteins. We performed the RaxML method under the WAG+I+G+F model with a nonparametric bootstrap. The tree is presented as a cladogram. (B) Schematic representation of Disa038 and phylogenetically related proteins. The betabaculovirus GPCR retained all the structures observed in the homologs including the signal peptide (gray), soluble fraction (black), and the transmembrane domains (TMDs, white). Plxy, *Plutella xylostella*; Psxu, *Papilio*

xuthus; Papo, *Papilio polytes*; MyseEV, *Mythimna separata entomopoxvirus 'L'*; and Dapl, *Danaus plexippus*.

3.5. GP64

The most striking aspect observed in the DisaGV genome was the presence of a *gp64* homolog gene, *disa118*. GP64 is the major envelope fusion protein (EFP) exclusively found in Group I alphabaculoviruses (G1-α) (Rohrmann, 2013). Both Group 2 alphabaculovirus (G2-α) and betabaculovirus share an analog to the GP64, called F protein, as the major BV EFP (Garry & Garry, 2008) which is probably the ancestral EFP in baculovirus (Jehle *et al.*, 2006a; Jehle *et al.*, 2006b). GP64 was acquired probably later by the ancestor of G1-α likely from an insect retrovirus-like element (Rohrmann & Karplus, 2001; Wang *et al.*, 2014) and is clearly related to the glycoprotein found in the genus *Thogotovirus* (from *Orthomyxoviridae*, an ssRNA negative-strand segmented virus family) (Morse *et al.*, 1992). Therefore, in attempt to understand both acquisition and evolution of *gp64* into the DisaGV genome, we performed a phylogenetic reconstruction of the gene. We found that DisaGV GP64 clustered with G1-α EFP, suggestive of a horizontal transfer from G1-α to betabaculovirus (Figure 6A). Disa-GP64 clustered with DekiNPV. Therefore, *gp64* gene acquisition probably caused an improvement in the ancestor of DisaGV as probably had happened to the G1-α. Taken together, these results suggest that the common ancestor of the G1-α acquired this gene once by HGT from some unknown source, which was later transferred to DisaGV or some related betabaculovirus ancestral. Alternatively, but less probably, the gene was firstly acquired by a DisaGV-related virus and later transferred to the common ancestor of G1-α. An adaptation of

disa118 to the G+C genome content of DisaGV was observed (Figure 6B) depicting that the gene acquisition is likely not recent (Monier *et al.*, 2007). Experimental analysis has shown that the incorporation of GP64 into the genome of Helicoverpa armigera nucleopolyhedrovirus, a G2 α-baculovirus, enhanced virus infectivity *in vivo* and *in vitro* (Shen *et al.*, 2012). GP64 and F protein can exploit either distinct (Westenberg *et al.*, 2007) or similar (Wang *et al.*, 2010) receptors to entry into the host. Therefore, *gp64* fixation has probably pervaded expansion in both fusion and biding virus activities (Liang *et al.*, 2005; Yu *et al.*, 2009) and could have functionally replaced the F protein in G1-α (Wang *et al.*, 2014). The evolutionary replacement hypothesis is reinforced by the fact that G1-α present a remnant F protein homolog in their genomes unable to compensate *gp64* loss and probably playing a role only on the virus pathogenicity (Lung *et al.*, 2003). Interestingly, despite the DisaGV genome codes for an F protein, large deletions were observed in several reads covering the gene, suggesting existence of viruses with deleted segments in the sequenced population (data not shown). This feature may indicate that the function of *f protein* has been replaced or complemented by *gp64* in DisaGV. Moreover, in our report, we analyzed the 150 nucleotides up-stream the predicted *gp64* ATG start codon from DisaGV to compare with annotations identified previously in G1 α-baculovirus *gp64* promoter region (Figure 6C). During viral *de novo* synthesis, *gp64* expression is regulated by transcription from both early and late promoters with negative and multiple positive regulatory elements (Blissard & Rohrmann, 1991). The *gp64* promoter region size was previously described to be around 140 bp (Chen *et al.*, 2013; Garrity *et al.*, 1997; Jarvis & Garcia, 1994). Concerning this region, we found 3 required elements GATA (-21, -89, and -104), 2 TATA Box-like (-35 and -76), 2 CACGTG-like (-38 and -61) sequences with mutation on the first C to A in both, and one TATA-box (-35)-associated CAGT (-38). TATA-

dependent activity and TATA-independent activity is mediated by RNA polymerase II in OpMNPV *gp64* (Kogan *et al.*, 1995). Two of the required GATA and CACGTG specifically bind to host transcription factors and activate transcription from the TATA-dependent *gp64* promoter (Kogan & Blissard, 1994; Kogan *et al.*, 1995). The presence of these conserved regulatory expression sequences in the promoter region of *disa-gp64* gene indicates that it must be transcribed and functional. We are currently analyzing whether *disa-gp64* is able to replace G1 α-baculovirus *gp64* gene.

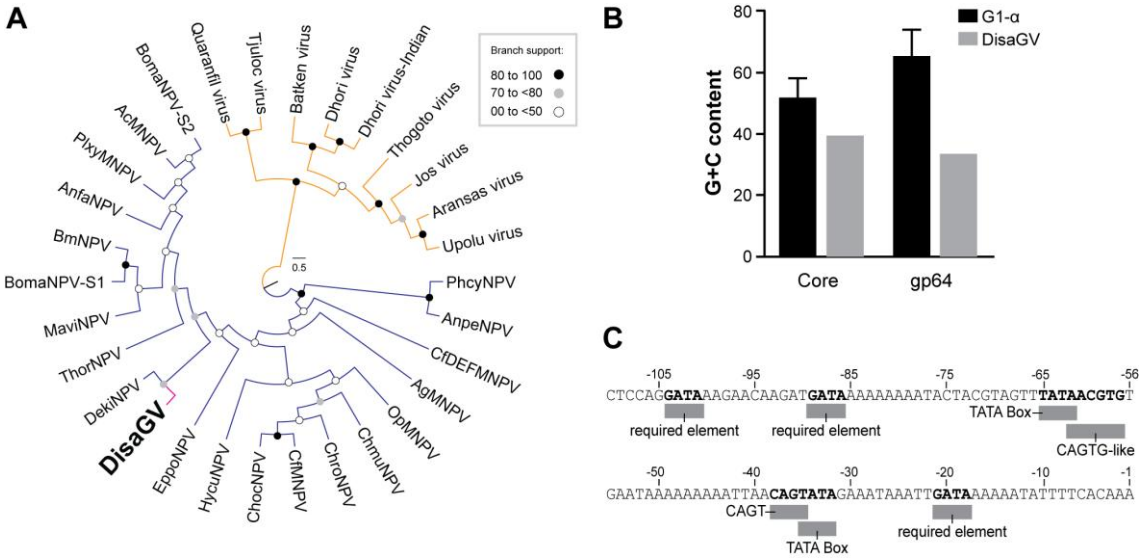

Figure 6. Phylogeny, G+C content, and the promoter region analyses of the betabaculovirus-encoded *gp64* homolog, *disa118*. (A) The DisaGV homolog is related to DekiNPV. The maximum likelihood (ML) tree was inferred using the predicted amino acid sequence of all the betabaculovirus GP64 (pink), several publicly available Group 1 alphabaculovirus genes (blue), and thogotovirus genes (orange). We performed the RaxML method under the WAG+I+G model with a nonparametric bootstrap for phylogeny reconstruction. Thogotoviruses root the tree that is presented here as a cladogram. (B) Comparison of the G+C content average for the third position of the

translational codon in the *gp64* genes from all Group 1 *Alphabaculovirus* (G1-α) and DisaGV. *disa118* underwent a gene adjustment for the low G+C content characteristic of betabaculoviruses when compared to the G1-α-derived genes. (C) Annotation of 110 bp long from the *disa118* promoter region. The elements and motifs were pictured based on previously published researches in alphabaculoviruses. We are presenting the required element GATA for *gp64* transcription, the TATA-boxes, and a CAGTG-like element.

4. Methods

4.1. Viral origin, confirmation, and electron microscopy

The DisaGV used in this study was obtained from infected larvae *D. saccharalis* collected in the state of Parana, Brazil in 2009. Transmission electron microscopy (TEM) of purified OBs and *granulin* gene amplicon sequence confirmed that the infection was due to a betabaculovirus.. The *granulin* amplification was performed with universal primers for the major OB protein gene as previously published (Lange *et al.*, 2004). The amplified fragment was purified from an agarose gel after electrophoresis with the GFX® kit (GE Healthcare) following the manufacturer`s instructions, Sanger sequencing reaction was performed with the BigDye kit (Applied Biosystems) and the sequence determined in an automated sequencer ABI Prism® 3100 Genetic Analyzer (Applied Biosystems). For transmission electron microscopy, a suspension of occlusion bodies extracted from larvae infected by DisaGV was prepared as described elsewhere (Ardisson-Araujo *et al.*, 2014a).

4.2. Sequencing system, assembly, and analysis of the DisaGV complete genome

DisaGV genomic DNA was sequenced with the 454 Genome Sequencer (GS) FLX™ Standard (Roche) at the Centro de Genômica de Alto Desempenho do Distrito Federal (Brasília, Brazil). The genome was assembled *de novo* using Geneious 7.0 (Kearse *et al.*, 2012) and confirmed using restriction enzyme digestion profile. The annotation was performed using Geneious 7.0 to identify the open reading frames (ORFs) that started with a methionine codon (ATG) encoding at least 50 amino acids and blastp to identify homologs. Specific primers were designed to amplify and sequence, by Sanger method, all regions in the genome with low coverage (< 10 x).

4.3. Phylogenetic analyses and genome comparison

For the *Baculoviridae* phylogeny, a MAFFT alignment (Katoh *et al.*, 2002) was carried out with the concatenated amino acid sequences predicted for the 37 baculovirus core genes. The hypothetical tree was inferred using the FastTree method (Liu *et al.*, 2011), implemented in Geneious. For the putative horizontal gene transfer (HGTs) events the same alignment method was used for Disa034, Disa038, Disa039 (G protein-encoding gene), and Disa118 (*gp64* homolog) and the hypothetical trees were inferred using the RaxML method with 100 repetitions of a non parametric bootstrap (Guindon *et al.*, 2010), implemented in Geneious, with the models WAG+I+G for GP64, WAG+I+G+F for Disa038, WAG+I+F for Disa039, and LG+I+G for Disa034 selected by Prottest 2.4 (Abascal *et al.*, 2005). The signal peptide and the transmembrane domains were predicted by both the SignalP 4.1 server (http://www.cbs.dtu.dk/services/SignalP/) and the TMHMM Sever v. 2.0 (http://www.cbs.dtu.dk/services/TMHMM/), respectively.

Moreover, the complete genome of DisaGV was compared with other betabaculovirus genomes through construction of syntenic maps with the Mauve program in the Genious 7.1.7 using default parameter settings.

5. Conclusion

After structural characterization, complete genome sequence, and phylogenetic analyses of the *Diatraea saccharalis*-infecting virus, we found that it is a novel species into the genus *Betabaculovirus*, called by *Diatraea saccharalis granulovirus* (DisaGV). The genome seemed to be closely related to CypoGV-related species and to present so far the smallest genome among other betabaculoviruses. Remarkably, we found in the genome both a GPCR-like and *gp64* gene. *gp64* used to be found solely in the group 1 alphabaculovirus genomes.

6. References

Abascal, F., Zardoya, R. & Posada, D. (2005). ProtTest: selection of best-fit models of protein evolution. *Bioinformatics* **21**, 2104-2105.

Ardisson-Araujo, D. M., de Melo, F. L., Andrade, M. D., Sihler, W., Bao, S. N., Ribeiro, B. M. & de Souza, M. L. (2014a). Genome sequence of Erinnyis ello granulovirus (ErelGV), a natural cassava hornworm pesticide and the first sequenced sphingid-infecting betabaculovirus. *BMC Genomics* **15**, 856.

Ardisson-Araujo, D. M., Melo, F. L., de Souza Andrade, M., Brancalhao, R. M., Bao, S. N. & Ribeiro, B. M. (2014b). Complete genome sequence of the first non-Asian isolate of Bombyx mori nucleopolyhedrovirus. *Virus Genes*.

Ardisson-Araujo, D. M., Sosa-Gómez, D. R., Melo, F. L., Báo, S. N. & Ribeiro, B. M. (2015). CHARACTERIZATION OF HELICOVERPA ZEA SINGLE NUCLEOPOLYHEDROVIRUS ISOLATED IN BRAZIL DURING THE FIRST OLD WORLD BOLLWORM (NOCTUIDAE: HELICOVERPA ARMIGERA) NATIONWIDE OUTBREAK. *Virus Reviews and Research* **20**, 4.

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

Chen, Y. R., Zhong, S., Fei, Z., Hashimoto, Y., Xiang, J. Z., Zhang, S. & Blissard, G. W. (2013). The Transcriptome of the Baculovirus Autographa californica Multiple Nucleopolyhedrovirus in Trichoplusia ni Cells. *J Virol* **87**, 6391-6405.

Craveiro, S. R., Inglis, P. W., Togawa, R. C., Grynberg, P., Melo, F. L., Ribeiro, Z. M., Ribeiro, B. M., Bao, S. N. & Castro, M. E. (2015). The genome sequence of Pseudoplusia includens single

nucleopolyhedrovirus and an analysis of p26 gene evolution in the baculoviruses. *BMC Genomics* **16**, 127.

Dinardo-Miranda, L. L. (2008). Pragas. In *Cana-de-açúcar*. Edited by L. L. Dinardo-Miranda, A. C. M. d. Vasconceles & M. G. d. A. Landell. Campinas: Instituto Agronômico.

Garrity, D. B., Chang, M. J. & Blissard, G. W. (1997). Late promoter selection in the baculovirus gp64 envelope fusion protein gene. *Virology* **231**, 167-181.

Garry, C. E. & Garry, R. F. (2008). Proteomics computational analyses suggest that baculovirus GP64 superfamily proteins are class III penetrenes. *Virol J* **5**, 28.

Guindon, S., Dufayard, J. F., Lefort, V., Anisimova, M., Hordijk, W. & Gascuel, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol* **59**, 307-321.

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

Ishimwe, E., Hodgson, J. J., Clem, R. J. & Passarelli, A. L. (2015a). Reaching the melting point: Degradative enzymes and protease inhibitors involved in baculovirus infection and dissemination. *Virology* **479-480**, 637-649.

Ishimwe, E., Hodgson, J. J. & Passarelli, A. L. (2015b). Expression of the Cydia pomonella granulovirus matrix metalloprotease enhances Autographa californica multiple nucleopolyhedrovirus virulence and can partially substitute for viral cathepsin. *Virology* **481**, 166-178.

Iyer, L. M., Balaji, S., Koonin, E. V. & Aravind, L. (2006). Evolutionary genomics of nucleo-cytoplasmic large DNA viruses. *Virus Res* **117**, 156-184.

Jarvis, D. L. & Garcia, A., Jr. (1994). Biosynthesis and processing of the Autographa californica nuclear polyhedrosis virus gp64 protein. *Virology* **205**, 300-313.

Jehle, J. A., Blissard, G. W., Bonning, B. C., Cory, J. S., Herniou, E. A., Rohrmann, G. F., Theilmann, D. A., Thiem, S. M. & Vlak, J. M. (2006a). On the classification and nomenclature of baculoviruses: a proposal for revision. *Arch Virol* **151**, 1257-1266.

Jehle, J. A., Lange, M., Wang, H., Hu, Z., Wang, Y. & Hauschild, R. (2006b). Molecular identification and phylogenetic analysis of baculoviruses from Lepidoptera. *Virology* **346**, 180-193.

Katoh, K., Misawa, K., Kuma, K. & Miyata, T. (2002). MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res* **30**, 3059-3066.

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Meintjes, P. & Drummond, A. (2012). Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28**, 1647-1649.

Kochman, K. (2014). Superfamily of G-protein coupled receptors (GPCRs)--extraordinary and outstanding success of evolution. *Postepy Hig Med Dosw (Online)* **68**, 1225-1237.

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

Kogan, P. H., Chen, X. & Blissard, G. W. (1995). Overlapping TATA-dependent and TATA-independent early promoter activities in the baculovirus gp64 envelope fusion protein gene. *J Virol* **69**, 1452-1461.

Krishnan, A., Almen, M. S., Fredriksson, R. & Schioth, H. B. (2012). The origin of GPCRs: identification of mammalian like Rhodopsin, Adhesion, Glutamate and Frizzled GPCRs in fungi. *PLoS One* **7**, e29817.

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

Lange, M., Wang, H., Zhihong, H. & Jehle, J. A. (2004). Towards a molecular identification and classification system of lepidopteran-specific baculoviruses. *Virology* **325**, 36-47.

Liang, C., Song, J. & Chen, X. (2005). The GP64 protein of Autographa californica multiple nucleopolyhedrovirus rescues Helicoverpa armigera nucleopolyhedrovirus transduction in mammalian cells. *J Gen Virol* **86**, 1629-1635.

Liu, K., Linder, C. R. & Warnow, T. (2011). RAxML and FastTree: comparing two methods for large-scale maximum likelihood phylogeny estimation. *PLoS One* **6**, e27731.

Lung, O. Y., Cruz-Alvarez, M. & Blissard, G. W. (2003). Ac23, an envelope fusion protein homolog in the baculovirus Autographa californica multicapsid nucleopolyhedrovirus, is a viral pathogenicity factor. *J Virol* **77**, 328-339.

Mahmoud, A. M., De Luna-Santillana, E. J. & Rodriguez-Perez, M. A. (2011). Parasitism by the endoparasitoid, Cotesia flavipes induces cellular immunosuppression and enhances susceptibility of the sugar cane borer, Diatraea saccharalis to Bacillus thuringiensis. *J Insect Sci* **11**, 119.

Monier, A., Claverie, J. M. & Ogata, H. (2007). Horizontal gene transfer and nucleotide compositional anomaly in large DNA viruses. *BMC Genomics* **8**, 456.

Morse, M. A., Marriott, A. C. & Nuttall, P. A. (1992). The glycoprotein of Thogoto virus (a tick-borne orthomyxo-like virus) is related to the baculovirus glycoprotein GP64. *Virology* **186**, 640-646.

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

Nijmeijer, S., Leurs, R., Smit, M. J. & Vischer, H. F. (2010). The Epstein-Barr virus-encoded G protein-coupled receptor BILF1 hetero-oligomerizes with human CXCR4, scavenges Galphai proteins, and constitutively impairs CXCR4 functioning. *J Biol Chem* **285**, 29632-29641.

Oliveira, J. V., Wolff, J. L., Garcia-Maruniak, A., Ribeiro, B. M., de Castro, M. E., de Souza, M. L., Moscardi, F., Maruniak, J. E. & Zanotto, P. M. (2006). Genome of the most widely used viral biopesticide: Anticarsia gemmatalis multiple nucleopolyhedrovirus. *J Gen Virol* **87**, 3233-3250.

Rohrmann, G. F. (2013). *Baculovirus Molecular Biology*. Bethesda (MD): National Center for Biotechnology Information (US). Available from: http://www.ncbi.nlm.nih.gov/books/NBK49500/.

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

Rossi, G. D., Salvador, G. & Consoli, F. L. (2014). THE PARASITOID, Cotesia flavipes (CAMERON) (HYMENOPTERA: BRACONIDAE), INFLUENCES FOOD CONSUMPTION AND UTILIZATION BY LARVAL Diatraea saccharalis (F.) (LEPIDOPTERA: CRAMBIDAE). *Arch Insect Biochem Physiol*.

Shen, S., Gan, Y., Wang, M., Hu, Z., Wang, H. & Deng, F. (2012). Incorporation of GP64 into Helicoverpa armigera nucleopolyhedrovirus enhances virus infectivity in vivo and in vitro. *J Gen Virol* **93**, 2705-2711.

Soccol, C. R., Vandenberghe, L. P., Medeiros, A. B., Karp, S. G., Buckeridge, M., Ramos, L. P., Pitarelo, A. P., Ferreira-Leitao, V., Gottschalk, L. M., Ferrara, M. A., da Silva Bon, E. P., de Moraes, L. M., Araujo Jde, A. & Torres, F. A. (2010). Bioethanol from lignocelluloses: Status and perspectives in Brazil. *Bioresour Technol* **101**, 4820-4825.

Wang, M., Wang, J., Yin, F., Tan, Y., Deng, F., Chen, X., Jehle, J. A., Vlak, J. M., Hu, Z. & Wang, H. (2014). Unraveling the entry mechanism of baculoviruses and its evolutionary implications. *J Virol* **88**, 2301-2311.

Wang, M., Yin, F., Shen, S., Tan, Y., Deng, F., Vlak, J. M., Hu, Z. & Wang, H. (2010). Partial functional rescue of Helicoverpa armigera single nucleocapsid nucleopolyhedrovirus infectivity by replacement of F protein with GP64 from Autographa californica multicapsid nucleopolyhedrovirus. *J Virol* **84**, 11505-11514.

Westenberg, M., Uijtdewilligen, P. & Vlak, J. M. (2007). Baculovirus envelope fusion proteins F and GP64 exploit distinct receptors to gain entry into cultured insect cells. *J Gen Virol* **88**, 3302-3306.

Wolff, J. L., Valicente, F. H., Martins, R., Oliveira, J. V. & Zanotto, P. M. (2008). Analysis of the genome of Spodoptera frugiperda nucleopolyhedrovirus (SfMNPV-19) and of the high genomic heterogeneity in group II nucleopolyhedroviruses. *J Gen Virol* **89**, 1202-1211.

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

Yu, I. L., Lin, Y. C., Robinson, J. H. & Lung, O. (2009). Transduction of vertebrate cells with Spodoptera exigua multiple nucleopolyhedrovirus F protein-pseudotyped gp64-null Autographa californica multiple nucleopolyhedrovirus. *J Gen Virol* **90**, 2282-2287.

Zanin, G. M., Santana, C. C., Bon, E. P., Giordano, R. C., de Moraes, F. F., Andrietta, S. R., de Carvalho Neto, C. C., Macedo, I. C., Fo, D. L., Ramos, L. P. & Fontana, J. D. (2000). Brazilian bioethanol program. *Appl Biochem Biotechnol* **84-86**, 1147-1161.

7. Supplementary material

Table S1. Gene composition and general features of the Diatraea saccharalis granulovirus (DisaGV) genome relative to other baculovirus genomes.

ORF	Name	Position	Size (bp)	Size (aa)	Transcriptional motifs	Orthologs - ORF number (identity)				
						AcMNPV	CypoGV	CrleGV	PiraGV	ChocGV
1	*granulin*	1 > 747	747	248	E, L	8 (55)	1 (93)	1 (91)	1 (81)	1 (91)
2		1084 < 728	357	118	E, L	-	2 (58)	2 (59)	2 (51)	2 (39)
3	*pk-1*	1065 > 1877	813	270	E, L	10 (38)	3 (54)	3 (58)	3 (61)	3 (61)
4		2112 < 1852	261	86	E	-	-	-	-	-
	hr1	1921 - 2134	214	-	-	-	-	-	-	-
5		2701 < 2132	570	189	E, L	-	4 (51)	4 (51)	4 (56)	5 (49)
6	*ie-1*	4218 < 2932	1287	428	E	147 (29)	7 (45)	6 (44)	6 (47)	7 (41)
7		4249 > 4818	570	189	E, L	146 (28)	8 (46)	7 (47)	7 (55)	8 (55)
8		5134 < 4838	297	98	E, L	145 (38)	9 (60)	8 (61)	8 (65)	9 (61)
9	*odv-e18*	5406 < 5140	267	88	E, L	143 (29)	14 (76)	13 (72)	14 (66)	12 (73)
10	*p49*	6790 < 5393	1398	465	E, L	142 (31)	15 (58)	14 (59)	15 (64)	13 (60)
11		7946 < 7368	579	192	E, L	-	16 (52)	15 (53)	-	-
12	*odv-e56*	9070 < 7943	1128	375	E	148 (46)	18 (70)	17 (68)	16 (68)	14 (72)
13		9096 > 9299	204	67	E	-	-	-	-	-
14	*pep1*	9757 < 9260	498	165	E, L	-	20 (57)	20 (66)	20 (53)	17 (70)
15		9837 > 10406	570	189	E, L	-	-	-	-	-
16	*pep/p10*	10453 > 11427	975	324	E, L	-	22 (64)	23 (63)	22 (60)	18 (64)
17	*pep2*	11439 > 11873	435	144	E, L	-	23 (71)	24 (70)	22 (70)	19 (66)
18		12838 < 11885	954	317	E, L	-	29 (36)	-	24 (27)	22 (29)
19		13103 < 12879	225	74	E, L	-	-	-	-	-
20		13528 > 14025	498	165	E, L	-	-	-	-	-
21	*gp41*	14904 < 14047	858	285	E, L	80 (32)	104 (67)	95 (69)	88 (71)	83 (66)
22		15445 < 14849	597	198	E, L	81 (47)	103 (73)	94 (71)	87 (71)	82 (70)
23		15740 < 15429	312	103	E	82 (26)	102 (47)	93 (43)	86 (47)	81 (45)
24	*vp91*	15718 > 17355	1638	545	E, L	83 (26)	101 (55)	92 (55)	89 (53)	80 (53)
25	*efp/f protein*	17418 > 19028	1611	536	E, L	23 (22)	31 (38)	30 (39)	26 (40)	23 (41)
26		19145 > 19630	486	161	E, L	-	-	-	-	-
27		19829 < 19587	243	80	E, L	-	-	-	-	-
28		19804 > 19989	186	61	-	-	-	-	-	-
29		20609 < 20007	603	200	E, L	-	33 (38)	32 (41)	28 (42)	24 (38)
30	*pif-3*	20637 > 21194	558	185	E, L	115 (39)	35 (53)	34 (48)	30 (52)	26 (50)
31	*odv-e66*	23470 < 21185	2286	761	L	46 (58)	37 (65)	35 (67)	45 (56)	27 (60)

32		23509 > 23820	312	103	E, L	-	39 (66)	36 (68)	31 (64)	28 (55)
33		24140 < 23856	285	94	E, L	-	-	-	-	-
34		25142 < 24210	933	310	E	-	-	-	-	-
35		25067 > 25219	153	50	E	-	-	-	-	-
36	*lef-2*	25278 > 25796	519	172	E	6 (27)	41 (48)	38 (49)	33 (54)	29 (51)
37		25783 > 26028	246	81	E, L	-	42 (33)	39 (37)	34 (48)	30 (41)
38		26098 > 27294	1197	398	E, L	-	-	-	-	-
39		27632 < 27291	342	113	E	-	-	-	-	-
40	*metalloproteinase*	28746 < 27634	1113	370	E, L	-	46 (36)	43 (40)	37 (40)	33 (39)
41	*p13*	28756 > 29556	801	266	E, L	-	47 (59)	44 (61)	38 (54)	34 (60)
	hr-2	29572 - 29929	358	-	-	-	-	-	-	-
42	*pif-2*	30025 > 31146	1122	373	E, L	22 (52)	48 (70)	45 (71)	40 (70)	35 (66)
43		31209 > 31592	384	127	E	-	-	-	-	-
44		31794 < 31597	198	65	E, L	-	49 (43)	46 (35)	41 (43)	-
45		31814 > 33667	1854	617	E, L	-	50 (46)	47 (51)	42 (31)	-
46		34271 < 33672	600	199	E, L	106 (39)	52 (68)	50 (71)	43 (72)	37 (72)
47		34283 > 34432	150	49	E, L	110 (24)	53 (67)	51 (63)	44 (76)	38 (81)
48	*v-ubq*	34775 < 34419	357	118	E	35 (79)	54 (82)	52 (84)	45 (82)	39 (85)
49	*odv-ec43*	34779 > 35816	1038	345	E, L	109 (32)	55 (56)	53 (58)	46 (69)	40 (65)
50		35822 > 36016	195	64	E, L	-	56 (47)	54 (56)	47 (45)	41 (58)
51	*39k/pp31*	36755 < 36018	738	245	E, L	36 (35)	57 (46)	52 (44)	48 (51)	55 (41)
52	*lef-11*	37026 < 36745	282	93	E, L	37 (27)	58 (66)	53 (64)	49 (59)	56 (56)
53	*p74*	39004 < 36950	2055	684	E, L	138 (42)	60 (60)	58 (61)	51 (60)	46 (58)
54		39497 < 39057	441	146	E, L	-	-	-	-	-
55	*acetyltransferase*	40081 < 39497	585	194	E, L	-	-	-	56 (66)	48 (58)
56		40513 < 40094	420	139	E, L	-	62 (70)	60 (58)	55 (47)	49 (86)
	hr-3	40146 - 40318	173	-	-	-	-	-	-	-
57	*p47*	40557 > 41711	1155	384	E, L	40 (42)	68 (65)	61 (65)	56 (66)	50 (66)
58	*bv-e31*	41746 > 42399	654	217	E, L	38 (42)	69 (76)	62 (75)	57 (76)	51 (72)
	hr-4	42416 - 43074	659	-	-	-	-	-	-	-
59		42631 < 42455	177	58	-	-	-	-	-	-
60	*p24*	43135 > 43626	492	163	E, L	129 (32)	71 (61)	63 (64)	58 (56)	52 (63)
61	*38.8k*	44072 < 43647	426	141	-	13 (30)	73 (35)	65 (40)	62 (45)	54 (37)
62	*lef-1*	44760 < 44053	708	235	E	14 (31)	74 (59)	66 (59)	60 (63)	55 (59)
63	*pif-1*	44770 > 46323	1554	517	E, L	119 (36)	75 (60)	67 (60)	61 (58)	56 (59)
64		46328 > 46690	363	120	E, L	-	70 (34)	-	-	-

65	*iap-3*	46790 > 47584	795	264	E, L	27 (32)	17 (54)	16 (48)	-	84 (50)
66		47630 > 47782	153	51	E, L	-	-	-	-	-
67		47800 > 48003	204	67	E, L	150 (22)	79 (35)	70 (35)	-	59 (38)
68	*lef-6*	48269 < 47985	285	94	E	28 (31)	80 (45)	71 (43)	65 (58)	60 (52)
69	*dbp*	49084 < 48287	798	265	E, L	25 (22)	81 (46)	72 (44)	66 (50)	61 (48)
70		49321 < 49103	219	72	E, L	-	82 (51)	73 (46)	70 (60)	62 (48)
71		49847 < 49263	585	194	E, L	-	82 (27)	73 (27)	67 (41)	63 (34)
72	*p48/p45*	49869 > 51041	1173	390	E, L	103 (33)	83 (70)	74 (68)	68 (73)	64 (69)
73		51068 > 51355	288	95	E, L	102 (26)	84 (50)	75 (49)	69 (50)	65 (40)
74		52171 < 51395	777	258	E, L	-	-	-	-	-
75	*odv-c42/p40*	52403 > 53563	1161	386	E, L	101 (22)	85 (59)	76 (58)	70 (57)	66 (55)
76	*p6.9*	53571 > 53756	186	61	E, L	-	-	-	-	-
77	*lef-5*	54493 < 53792	702	234	E, L	99 (42)	87 (68)	78 (68)	72 (71)	68 (66)
78	*38 k*	54443 > 55354	912	303	E, L	98 (39)	88 (60)	79 (59)	73 (73)	69 (66)
79	*dut*	55338 > 55808	471	156	E, L	-	-	-	-	-
80		55805 > 56179	375	124	-	-	-	-	-	-
81	*odv-e28/pif-4*	56684 < 56199	486	161	E, L	96 (35)	89 (62)	80 (64)	74 (61)	70 (55)
82	*helicase-1*	56668 > 60051	3384	1127	E, L	95 (26)	90 (52)	81 (52)	75 (57)	71 (54)
83	*odv-e25*	60707 < 60069	639	212	E, L	94 (37)	91 (69)	82 (69)	76 (68)	72 (70)
84	*p18*	61209 < 60727	483	160	E, L	93 (33)	92 (44)	83 (40)	77 (49)	73 (46)
85	*sox/p33*	61224 > 61979	756	251	E, L	92 (36)	93 (66)	84 (64)	78 (67)	74 (66)
86	*lef-4*	63298 < 61976	1323	440	E, L	90 (32)	95 (54)	86 (52)	80 (57)	75 (55)
87	*vp39*	63312 > 64166	855	284	E, L	89 (33)	96 (60)	87 (62)	81 (63)	76 (61)
88	*odv-ec27*	64217 > 65029	813	270	L	144 (31)	97 (61)	88 (61)	82 (64)	77 (55)
	hr5	65061 - 65488	428	-	-	-	-	-	-	-
89		65127 > 65417	291	96	E	-	-	-	-	-
90		66499 < 65465	1035	344	E, L	-	99 (35)	90 (35)	83 (36)	78 (37)
91		66528 > 66722	195	64	E, L	-	100 (50)	91 (51)	84 (50)	79 (60)
92		66729 > 66995	267	88	E, L	78 (42)	105 (41)	96 (46)	89 (45)	85 (48)
93	*vlf-1*	66940 > 68055	1116	371	E, L	77 (34)	106 (73)	97 (74)	90 (77)	86 (67)
94		68079 > 68330	252	83	E, L	76 (26)	107 (67)	98 (65)	91 (70)	88 (68)
95		68345 > 68800	456	151	E, L	75 (28)	108 (57)	99 (58)	92 (63)	89 (63)
96		69174 < 68830	345	114	E, L	-	110 (26)	100 (23)	-	-
97	*dna pol*	72280 < 69203	3078	1025	E	65 (33)	111(65)	101(66)	93 (68)	90 (66)
98	*desmoplakin*	72255 > 73895	1641	546	E, L	66 (27)	112 (31)	102 (32)	98 (32)	91 (33)
	hr-6	73837 - 74251	415	-	-	-	-	-	-	-

99	*lef-3*	75316 < 74249	1068	355	E, L	67 (24)	113 (35)	103 (38)	95 (45)	92 (43)
100	*odv-nc42*	75282 > 75668	387	128	E, L	68 (34)	114 (65)	104 (57)	96 (65)	93 (62)
101		75849 < 75661	189	62	E	-	-	-	-	-
102		75766 > 76248	483	160	E, L	-	115 (38)	105 (33)	97 (39)	94 (33)
103	*iap-5*	76307 > 77101	795	264	E	-	116 (63)	106 (60)	98 (58)	95 (58)
104	*lef-9*	77106 > 78584	1479	492	L	62 (53)	117 (73)	107 (72)	99 (75)	96 (73)
105	*fp25k*	78590 > 79030	441	146	E, L	61 (36)	118 (68)	108 (66)	100 (72)	97 (67)
106		80190 < 79057	1134	377	E	-	-	-	-	-
107		80444 < 80193	252	83	E	-	-	-	-	-
108		82066 < 80555	1512	503	E	-	-	-	-	-
109	*dna ligase*	83743 < 82106	1638	545	E, L	-	120 (59)	110 (59)	102 (59)	99 (58)
110		83942 < 83745	198	65	E, L	-	-	-	-	-
111		84027 > 84341	315	104	E, L	-	124 (55)	114 (56)	106 (49)	103 (55)
112		84499 < 84317	183	60	E	-	-	-	-	-
113	*alk-exo*	84414 > 85610	1197	398	E, L	133 (33)	125 (53)	115(53)	107 (64)	104 (56)
114	*helicase-2*	85513 > 86805	1293	430	E, L	-	126 (59)	116 (54)	108 (59)	105 (52)
115	*rr1*	88666 < 86849	1818	605	E	-	127 (54)	-	-	-
116	*rr2a*	88765 > 89892	1128	375	L	-	128 (59)	-	-	-
117		89960 < 89808	153	50	E	-	-	-	-	-
118	*gp64*	89952 > 91469	1518	505	E, L	128 (74)	-	-	-	-
119	*lef-8*	93976 < 91472	2505	834	E, L	50 (48)	131 (69)	119 (67)	110 (70)	107 (68)
120		94000 > 94404	405	134	E, L	53 (36)	134 (67)	121 (69)	113 (68)	109 (63)
	hr-7	94399 - 94911	513	-	-	-	-	-	-	-
121		95772 < 94963	810	269	E, L	-	135 (41)	122 (29)	114 (38)	110 (31)
122		96139 < 95942	198	65	E, L	-	136 (41)	123 (44)	115 (50)	111 (44)
123	*lef-10*	96120 > 96353	234	77	E, L	53a (38)	137 (52)	124 (56)	120 (61)	112 (55)
124	*vp1054*	96214 > 97197	984	327	E, L	54 (30)	138 (59)	125 (58)	116 (66)	113 (59)
125	*me53*	97422 > 98366	945	314	E	139 (27)	143 (52)	129 (47)	125 (52)	116 (46)

Note: Position, transcriptional orientation and length (bp and aa) of 125 putative ORFs of the DisaGV genome. The ORFs were compared with their respective homologs in AcMNPV and 4 betabaculoviruses in terms of corresponding ORF number and amino acid identity (ID %). DisaGV unique ORFs are shown in red, betabaculovirus-specific ORFs in green, ORFs conserved in all baculovirus genomes (core genes) in blue. The conserved early (E; TATAW, TATAWTW e/ou TATAWAW) and late (L; A/T/GTAAG) transcriptional motifs within 450 bp upstream each putative ORF are also shown.

Table S2. Species used in this paper for reconstruction of the baculovirus phylogeny in the FIG. 3A. The species from the genera *Alphabaculovirus* (dark blue), *Betabaculovirus* (pink), *Gammabaculovirus* (orange), and *Deltabaculovirus* (light blue) are presented here together with the abbreviation used in the main text, the host family where the virus was isolated from, and the Genbank accession number as well.

Species	Abbreviation	Host family	Accession
Adoxophyes honmai nucleopolyhedrovirus	AdhoNPV	Tortricidae	AP006270
Adoxophyes orana nucleopolyhedrovirus	AdorNPV	Tortricidae	EU591746
Agrotis ipsilon multiple nucleopolyhedrovirus strain illinois	AgipMNPV	Noctuidae	EU839994
Agrotis segetum nucleopolyhedrovirus	AgseNPV	Noctuidae	DQ123841
Apocheima cinerarium nucleopolyhedrovirus	ApciNPV	Geometridae	FJ914221
Buzura suppressaria nucleopolyhedrovirus	BusuNPV	Geometridae	KF611977
Chrysodeixis chalcites nucleopolyhedrovirus	ChchNPV	Noctuidae	AY864330
Clanis bilineata nucleopolyhedrovirus	ClbiNPV	Sphingidae	DQ504428
Ectropis obliqua nucleopolyhedrovirus strain A1	EcobNPV-A1	Geometridae	DQ837165
Euproctis pseudoconspersa nucleopolyhedrovirus	EupsNPV	Lymantriidae	FJ227128
Helicoverpa armigera multiple nucleopolyhedrovirus	HaMNPV	Noctuidae	EU730893
Helicoverpa armigera nucleopolyhedrovirus C1	HaNPV-C1	Noctuidae	AF303045
Helicoverpa zea single nucleopolyhedrovirus USA	HzSNPV-USA	Noctuidae	AF334030
Hemileuca sp. nucleopolyhedrovirus	HespNPV	Saturniidae	KF158713
Lambdina fiscellaria nucleopolyhedrovirus	LafiNPV	Geometriidae	KP752043
Leucania separata nuclear polyhedrovirus strain AH1	LeseNPV	Noctuidae	AY394490
Lymantria díspar multiple nucleopolyhedrovirus	LdMNPV	Lymantriidae	AF081810
Lymantria xylina multiple nucleopolyhedrovirus	LyxyMNPV	Lymantriidae	GQ202541
Mamestra brassicae multiple nucleopolyhedrovirus strain Chb1	MbMNPV-CHb1	Noctuidae	JX138237

Mamestra configurata nucleopolyhedrovirus-A strain 90/2	MacoNPV-A 90/2	Noctuidae	U59461
Mamestra configurata nucleopolyhedrovirus B	MacoNPV-B	Noctuidae	AY126275
Orgyia leucostigma nucleopolyhedrovirus isolate CFS-77	OrleNPV	Lymantriidae	EU309041
Peridroma sp. nucleopolyhedrovirus	PespNPV	Noctuidae	KM009991
Perigonia lusca single nucleopolyhedrovirus	PeluSNPV	Sphigidae	KM596836
Pseudoplusia includens single nucleopolyhedrovirus IE	PsinSNPV	Noctuidae	KJ631622
Spodoptera exigua nucleopolyhedrovirus	SeMNPV	Noctuidae	AF169823
Spodoptera frugiperda multiple nucleopolyhedrovirus isolate 19	SfMNPV-19	Noctuidae	EU258200
Spodoptera litoralis nucleopolyhedrovirus isolate AN1956	SpliNPV-1956	Noctuidae	JX454574
Spodoptera litura nucleopolyhedrovirus G2	SpliNPV-G2	Noctuidae	AF325155
Spodoptera litura nucleopolyhedrovirus II	SpliNPV-II	Noctuidae	EU780426
Sucra jujuba nucleopolyhedrovirus	SujuNPV	Geometridae	KJ676450
Trichoplusia ni single nucleopolyhedrovirus	TnSNPV	Noctuidae	DQ017380
Autographa californica nucleopolyhedrovirus clone C6	AcMNPV-C6	Noctuidae	L22858
Anticarsia gemmatalis nucleopolyhedrovirus	AgMNPV	Noctuidae	DQ813662
Antheraea pernyi nucleopolyhedrovirus isolate L2	AnpeNPV-L2	Saturniidae	EF207986
Bombyx mori nucleopolyhedrovirus strain T3	BmNPV-T3	Bombycidae	L33180
Bombyx mandarina nucleopolyhedrovirus S2	BomaNPV-S2	Bombycidae	JQ071499
Choristoneura fumiferana defective multiple nucleopolyhedrovirus	CfDEFMNPV	Tortricidae	AY327402
Choristoneura fumiferana multiple nucleopolyhedrovirus	CfMNPV	Tortricidae	AF512031
Choristoneura murinana nucleopolyhedrovirus	ChmuNPV	Tortricidae	KF894742
Choristoneura occidentalis nucleopolyhedrovirus	ChocNPV	Tortricidae	KC961303
Choristoneura rosaceana nucleopolyhedrovirus	ChroNPV	Tortricidae	KC961304
Condylorrhiza vestigialis multiple nucleopolyhedrovirus	CoveMNPV	Crambidae	KJ631623

Dendrolimus kikuchii nucleopolyhedrovirus	DekiNPV	Lasiocampidae	JX193905
Epiphyas postvittana nucleopolyhedrovirus	EppoNPV	Tortricidae	AY043265
Hyphantria cunea nucleopolyhedrovirus	HycuNPV	Arctiidae	AP009046
Maruca vitrata multiple nucleopolyhedrovirus	MaviMNPV	Crambidae	EF125867
Orgyia pseudotsugata multiple nucleopolyhedrovirus	OpMNPV	Lymantriidae	U75930
Philosamia cynthia ricini nucleopolyhedrovirus	PhcyNPV	Saturniidae	JX404026
Plutella xylostella multiple nucleopolyhedrovirus isolate CL3	PlxyMNPV	Plutellidae	DQ457003
Rachiplusia ou multiple nucleopolyhedrovirus	RoMNPV	Noctuidae	AY145471
Thysanoplusia orichalcea nucleopolyhedrovirus	ThorNPV	Noctuidae	JX467702
Adoxophyes orana granulovirus	AdorGV	Tortricidae	AF547984
Agrotis segetum granulovirus-L1	AgseGV-L1	Noctuidae	KC994902
Choristoneura occidentalis granulovirus	ChocGV	Tortricidae	DQ333351
Clostera anastomosis granulovirus	CaLGV	Notodontidae	KC179784
Clostera anachoreta granulovirus	ClanGV	Notodontidae	HQ116624
Clostera anastomosis granulovirus Strain B	ClanGV-B	Notodontidae	KR091910
Cryptophlebia leucotreta granulovirus isolate CV3	CrleGV	Tortricidae	AY229987
Cydia pomonella granulovirus	CpGV	Tortricidae	U53466
Diatraea saccharalis granulovirus	**DisaGV**	**Crambidae**	**KP296186**
Epinotia aporema granulovirus	EpapGV	Tortricidae	JN408834
Erinnyis ello granulovirus	ErelGV	Sphingidae	KJ406702
Helicoverpa armigera granulovirus	HaGV	Noctuidae	EU255577
Phthorimaea operculella granulovirus	PhopGV	Gelechiidae	AF499596
Pieris rapae granulovirus E3	PiraGV-E3	Pieridae	GU111736
Plutella xylostella granulovirus	PlxyGV	Plutellidae	AF270937

Pseudaletia unipuncta granulovirus	PsunGV-Hawaiin	Noctuidae	EU678671
Spodoptera frugiperda granulovirus	SpfrGV	Noctuidae	KM371112
Spodoptera litura granulovirus isolate K1	SpliGV	Noctuidae	DQ288858
Xestia c-nigrum granulovirus	XcGV	Noctuidae	AF162221
Neodiprion sertifer nucleopolyhedrovirus	NeseNPV	Diprionidae	AY430810
Neodiprion lecontei nucleopolyhedrovirus	NeleNPV	Diprionidae	AY349019
Neodiprion abietis nucleopolyhedrovirus	NeabNPV	Diprionidae	DQ317692
Culex nigripalpus nucleopolyhedrovirus	CuniNPV	Culicidae	AF403738

Capítulo 7. A betabaculovirus-enconded *gp64* homolog is a functional envelope fusion protein

1. SUMMARY

The envelope fusion protein GP64 is a hallmark of group I alphabaculoviruses. However, the Diatraea saccharalis granulovirus genome sequence revealed the first betabaculovirus species harboring a *gp64* homolog (*disa118*). In this work, we have shown that this homolog is a functional envelope fusion protein and could enable infection and fusogenic abilities of a *gp64*-null prototype baculovirus. Therefore, GP64 may complement or may be in the process of replacing F protein activity in this virus lineage.

Este capítulo não foi publicado. **Ardisson-Araujo, D. M., Melo, F. L., Clem, R. J., Wolff, J. L., & Ribeiro, B. M. (2014).** A betabaculovirus-enconded *gp64* homolog is a functional envelope fusion protein

2. MAIN TEXT

The *Baculoviridae* is a family of insect viruses with double-stranded DNA genomes. It is currently divided into four genera, two of which, *Alphabaculovirus* and *Betabaculovirus*, contain members that are infective to the larval stages of moths and butterflies. During a complete infection cycle, viruses from both genera produce two virion phenotypes, (1) the occlusion-derived virus (ODV) which is surrounded by a crystalline protein matrix, the occlusion body (OB), and is responsible for the inter-host

oral primary infection and (2) the budded virion (BV), responsible for intra-host systemic infection (Rohrmann, 2013). GP64 is the major envelope fusion protein (EFP) found in the BVs of all group I alphabaculoviruses (G1-α) (Rohrmann, 2013). Other baculoviruses including those from group II alphabaculovirus (G2-α) and betabaculovirus share a GP64 analog called F protein as the major BV EFP (Garry & Garry, 2008).

The betabaculovirus Diatraea saccharalis granulovirus (DisaGV) was isolated from one of the most devastating insect pest of sugarcane and other cultures in Brazil. After complete genome sequencing (Genbank accession number: KP296186), a *gp64* homolog, *disa118* was found (unpublished data). *disa118* clustered with genes from alphabaculovirus group I instead of orthomyxovirus homologs, which confirms that *gp64* was acquired once by alphabaculovirus and then transferred to either DisaGV or a related ancestor (unpublished data). GP64 is a class III integral membrane glycoprotein (Garry & Garry, 2008) that plays essential roles in host cell receptor binding (Hefferon *et al.*, 1999), low-pH-triggered viral membrane fusion, (Kingsley *et al.*, 1999) and systemic infection of the host insect (Monsma *et al.*, 1996). Here, we investigated whether the *gp64* (*disa118*) homolog found in DisaGV is a functional EFP.

To examine the function of the DisaGV gp64 homolog, we generated a *gp64*-null Autographa californica multiple nucleopolyhedrovirus (Ac-Δ*gp64*-PG) bacmid pseudotyped with the *disa118* gene (hereby called *disa-gp64*). The pseudotyped virus Ac-REP-*disa-gp64*-PG was able to infect and spread upon transfection into *Spodoptera frugiperda* cell line 9 (Sf9) (FIG. 1A). To confirm that infectious BVs were being produced after transfection, we transferred the supernatants from transfection to healthy

Sf9 cell cultures. Ac-REP-*disa-gp64*-PG was able to cause infection (FIG. 1B). However, the efficiency was lower than the control viruses in a controlled infection assay (triplicate infection with MOI 5 and 6 h rocking, FIG 1C). In a previous study, a *gp64*-null virus expressing the EFP of the vesicular stomatitis virus G was able to produce infection, replicate, and propagate in Sf9 cells despite the cell-to-cell propagation being delayed in comparison to the parental virus (Mangor *et al.*, 2001). Interestingly, even with a pairwise identity of 73.2 % between Ac-GP64 and Disa-GP64, a monoclonal antibody against Ac-GP64 was unable to recognize Disa-GP64. However, a polyclonal antibody raised against Anticarsia gemmatalis multiple nucleopolyhedrovirus GP64 lacking both the signal peptide and the transmembrane domain recognized both Disa-GP64 and Ac-GP64 (FIG. 1D). We also carried out a fusogenic activity assay to verify whether the betabaculovirus glycoprotein could mediate low-pH-triggered membrane fusion. We found that cells infected with both vAc-REP-*disa-gp64*-PG and vAc-REP-*ac-gp64*-PG mediated membrane fusion and syncytium formation when exposed to low pH (FIG 2A and B, respectively). The efficiency of syncytium formation was apparently much lower when compared to the positive control. Moreover, no syncytium formation was observed when the cells were mock infected and treated with low pH (data not shown).

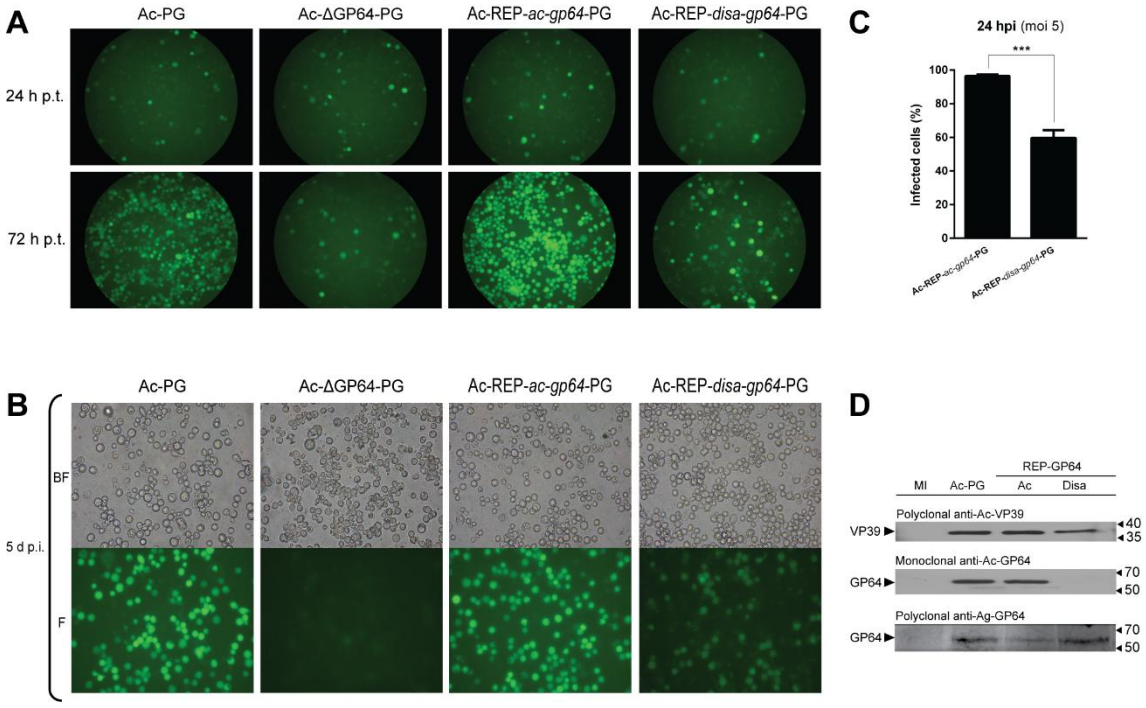

FIG. 1. Disa-GP64 is a functional envelope fusion protein. (A) Transfection assay of Ac-PG (positive control), Ac-ΔGP64-PG (negative control), Ac-REP-*ac-gp64*-PG (repaired virus), and Ac-REP-*disa-gp64*-PG (pseudotyped virus). 1 µg of DNA from each virus was transfected into Sf9 cells. The cells were photographed at 24 and 72 h p.i. (B) Ac-REP-*disa-gp64*-PG transfection supernatant is infective to Sf9 cells. At 5 days post-transfection, clarified supernatants were used to infect Sf9 cells. The cells were photographed at 5 days post-infection. (C) The infection efficiency of the pseudotyped Ac-REP-*disa-gp64*-PG was reduced when compared to the repaired Ac-REP-*ac-gp64*-PG. Cells were infected with MOI of 5 (determined by end-point dilution) and photographed at 24 hpi. (D) A monoclonal anti-Ac-GP64 does not recognize Disa-GP64 when expressed by recombinant AcMNPV but a polyclonal anti-Ag-GP64 does. Anti-Ac-VP39 antibody was used as a baculovirus infection control. Cells were mock infected or infected with (i) Ac-PG, (ii) Ac-REP-*ac-gp64*-PG (Ac), or (iii) Ac-REP-*disa-gp64*-PG (Disa) at an MOI of 5 for 72 hpi. Cells were harvested, and

the total proteins were extracted, resolved on SDS-12% PAGE gels, and analyzed by immunoblotting with polyclonal anti-Ac-VP39, monoclonal anti-Ac-GP64, or polyclonal anti-Ag-GP64 antibody.

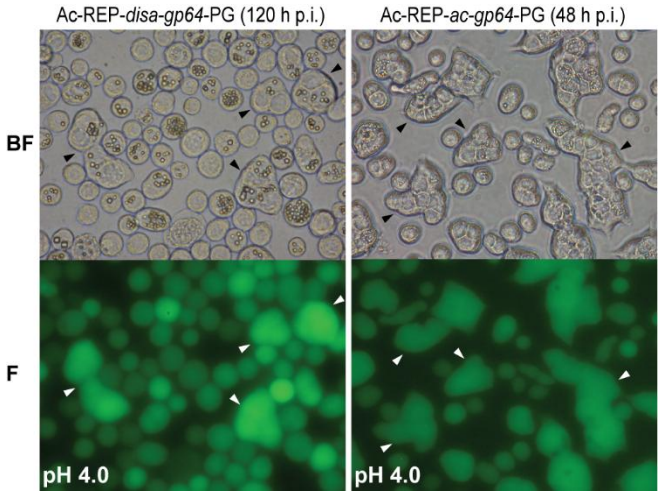

FIG.2. Syncitium formation mediated by recombinant baculovirus infections. Sf9 cells were infected with either AcRep-*Disa-gp64*-PG or AcRep-*Ac-gp64*-PG at MOI of 1. The infected cells were then incubated with low pH TC100 media (pH 4.0) for 10 min at 48 or 120 hpi, as indicated. After 10 min the media was replaced by media at pH 6.0. Syncitium formation was observed and photographed at 4 h after treatment. Multinucleated cells are indicated by arrow heads. The absence of OBs is due to the different time pos-infection used for the repaired virus.

To understand this reduction in virus infectivity, spread, and syncytium formation efficiency, we mapped functionally important amino acid residues in Disa-GP64 based on previous reports and protein alignment (Fig. 3). Two main regions were analyzed which included the signal peptide (SP) and the ectodomain (ED, region between the SP

and the transmembrane domain, TMD). By MAFFT alignment (Katoh *et al.*, 2002), we found that GP64 SPs across baculovirus species are variable *per se*, with a pairwise identity of 39.8% (Fig. 3A), which is not an exclusive feature in GP64 alone. Other baculoviral envelope proteins and secreted enzymes present highly variable SP sequences as well (*e.g. per os* infectivity factors and EGT) (data not shown). These amino acid substitutions could be related to host adaptation and might be responsible for the efficiency reduction displayed by the pseudotyped virus since DisaGV and AcMNPV were found infecting caterpillars from different lepidopteran families *i.e.* Crambidae and Noctuidae, respectively. On the other hand, the ED has been shown to present important regions for the functions of GP64 (Katou *et al.*, 2010; Li & Blissard, 2009; Zhou & Blissard, 2008). Using the same alignment method cited above, we found that most of the previously mapped ED regions and sites are highly conserved in Disa-GP64 such as intra-molecular disulfide bonds, which are critical in membrane fusion (Li & Blissard, 2010) (not shown). However, out of four glycosylation sites identified in Ac-GP64 ED (N198, N355, N385, and N426) and conserved in all other G1-α GP64 orthologs, three are maintained in Disa-GP64; only N355 underwent a substitution (Fig. 3B). Cell surface expression, assembly into infectious BV, and fusogenic activity do not require N-linked oligosaccharide processing; however, the removal of one or more N-glycosylation sites in Ac-GP64 impairs binding of budded virus to the cell, indicating that this modification is necessary for optimal GP64 function (Jarvis & Garcia, 1994; Jarvis *et al.*, 1998). Interestingly, both the production of infectious BV and the fusion activity were reduced when glycosylation of GP64 was inhibited in Bombyx mori nucleopolyhedrovirus (Rahman & Gopinathan, 2003).

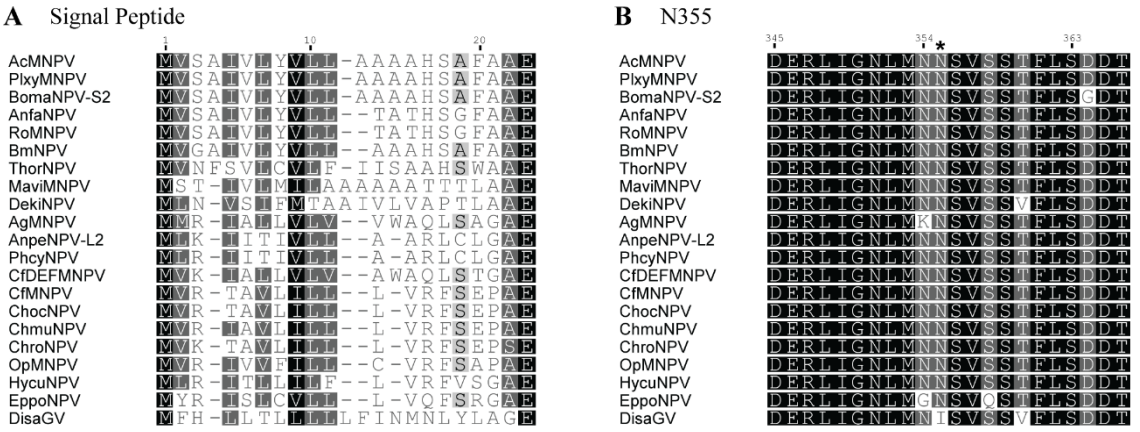

FIG. 3. Aligned regions of GP64 homologs from group I alphabaculoviruses and DisaGV. (A) Signal peptide region alignment. The last residue shown (glutamate) is the predicted beginning of the soluble portion of the protein. (B) Alignment of part of the soluble portion revealing the substitution in DisaGV from N355 to I355 when compared to the other alphabaculovirus species (asterisk). This residue has been experimentally shown to be a N-linked glycosylation site in AcMNPV. By the MAFFT alignment method, strictly conserved amino acid residues are shown in black boxes and partially conserved residues in grey boxes.

The main question here is why has *gp64* been fixed into DisaGV? Fixation of *gp64* is responsible for improvement of both fusion and biding activities (Liang *et al.*, 2005; Shen *et al.*, 2012; Yu *et al.*, 2009), and possibly led to replacement of F protein in G1-α (Wang *et al.*, 2014). In fact, G1-α viruses also contain a remnant F protein homolog in their genomes that is unable to compensate for *gp64* loss (14, 15), and that plays a role in virus pathogenicity (Lung *et al.*, 2003). Previous experimental analysis has shown that the incorporation of GP64 into a G2-α enhanced virus infectivity *in vivo* and *in vitro* (Shen *et al.*, 2012). Since *D. saccharalis* is an insect borer during the larval stage and presents a very short time of virus exposure between the egg hatching and the insect

penetration into the host plant which includes sugar cane, rice, and other monocots, it is reasonable to propose that a novel gene acquisition occurred that allowed the virus to improve its spread within the host and more effectively establish infection.

In summary, GP64 of DisaGV is a functional EFP that is able to pseudotype a *gp64*-null AcMNPV, although with a lower efficiency in spreading the infection and in fusogenic activity. The lack of one conserved glycosylation site and the possible adaptation to a different lepidopteran-family cell machinery could explain this reduction. We are constructing different mutants of *disa-gp64* to test those hypotheses. Importantly, in submitted work describing the DisaGV genome, we found several early transcriptional motifs upstream the *gp64* start codon; however, it is not clear whether DisaGV express the *gp64* homolog and uses it as a functional EPF. We can only speculate that GP64 could complement or may even be in the process of replacing F protein activity in this betabaculovirus lineage.

5. REFERENCES

Garry, C. E. & Garry, R. F. (2008). Proteomics computational analyses suggest that baculovirus GP64 superfamily proteins are class III penetrenes. *Virol J* **5**, 28.

Hefferon, K. L., Oomens, A. G., Monsma, S. A., Finnerty, C. M. & Blissard, G. W. (1999). Host cell receptor binding by baculovirus GP64 and kinetics of virion entry. *Virology* **258**, 455-468.

Jarvis, D. L. & Garcia, A., Jr. (1994). Biosynthesis and processing of the Autographa californica nuclear polyhedrosis virus gp64 protein. *Virology* **205**, 300-313.

Jarvis, D. L., Wills, L., Burow, G. & Bohlmeyer, D. A. (1998). Mutational analysis of the N-linked glycans on Autographa californica nucleopolyhedrovirus gp64. *J Virol* **72**, 9459-9469.

Katoh, K., Misawa, K., Kuma, K. & Miyata, T. (2002). MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res* **30**, 3059-3066.

Katou, Y., Yamada, H., Ikeda, M. & Kobayashi, M. (2010). A single amino acid substitution modulates low-pH-triggered membrane fusion of GP64 protein in Autographa californica and Bombyx mori nucleopolyhedroviruses. *Virology* **404**, 204-214.

Kingsley, D. H., Behbahani, A., Rashtian, A., Blissard, G. W. & Zimmerberg, J. (1999). A discrete stage of baculovirus GP64-mediated membrane fusion. *Mol Biol Cell* **10**, 4191-4200.

Li, Z. & Blissard, G. W. (2009). The pre-transmembrane domain of the Autographa californica multicapsid nucleopolyhedrovirus GP64 protein is critical for membrane fusion and virus infectivity. *J Virol* **83**, 10993-11004.

Li, Z. & Blissard, G. W. (2010). Baculovirus GP64 disulfide bonds: the intermolecular disulfide bond of Autographa californica multicapsid nucleopolyhedrovirus GP64 is not essential for membrane fusion and virion budding. *J Virol* **84**, 8584-8595.

Liang, C., Song, J. & Chen, X. (2005). The GP64 protein of Autographa californica multiple nucleopolyhedrovirus rescues Helicoverpa armigera nucleopolyhedrovirus transduction in mammalian cells. *J Gen Virol* **86**, 1629-1635.

Lung, O. Y., Cruz-Alvarez, M. & Blissard, G. W. (2003). Ac23, an envelope fusion protein homolog in the baculovirus Autographa californica multicapsid nucleopolyhedrovirus, is a viral pathogenicity factor. *J Virol* **77**, 328-339.

Mangor, J. T., Monsma, S. A., Johnson, M. C. & Blissard, G. W. (2001). A GP64-null baculovirus pseudotyped with vesicular stomatitis virus G protein. *J Virol* **75**, 2544-2556.

Monsma, S. A., Oomens, A. G. & Blissard, G. W. (1996). The GP64 envelope fusion protein is an essential baculovirus protein required for cell-to-cell transmission of infection. *J Virol* **70**, 4607-4616.

Rahman, M. M. & Gopinathan, K. P. (2003). Characterization of the gene encoding the envelope fusion glycoprotein GP64 from Bombyx mori nucleopolyhedrovirus. *Virus Res* **94**, 45-57.

Rohrmann, G. F. (2013). *Baculovirus Molecular Biology*. Bethesda (MD): National Center for Biotechnology Information (US). Available from: http://www.ncbi.nlm.nih.gov/books/NBK49500/.

Shen, S., Gan, Y., Wang, M., Hu, Z., Wang, H. & Deng, F. (2012). Incorporation of GP64 into Helicoverpa armigera nucleopolyhedrovirus enhances virus infectivity in vivo and in vitro. *J Gen Virol* **93**, 2705-2711.

Wang, M., Wang, J., Yin, F., Tan, Y., Deng, F., Chen, X., Jehle, J. A., Vlak, J. M., Hu, Z. & Wang, H. (2014). Unraveling the entry mechanism of baculoviruses and its evolutionary implications. *J Virol* **88**, 2301-2311.

Yu, I. L., Lin, Y. C., Robinson, J. H. & Lung, O. (2009). Transduction of vertebrate cells with Spodoptera exigua multiple nucleopolyhedrovirus F protein-pseudotyped gp64-null Autographa californica multiple nucleopolyhedrovirus. *J Gen Virol* **90**, 2282-2287.

Zhou, J. & Blissard, G. W. (2008). Identification of a GP64 subdomain involved in receptor binding by budded virions of the baculovirus Autographica californica multicapsid nucleopolyhedrovirus. *J Virol* **82**, 4449-4460.

Capítulo 8. Genome sequence of Perigonia lusca single nucleopolyhedrovirus (PeluSNPV): insights on the evolution of a nucleotide metabolism enzyme in the family *Baculoviridae*

1. Abstract

The genome of a novel group II alphabaculovirus, Perigonia lusca single nucleopolyhedrovirus (PeluSNPV), was sequenced and shown to contain 132,831 bp with 145 putative ORFs (open reading frames) encoding polypeptides with at least 50 amino acid residues. Among the 145 ORFs, 18 were found to be unique and, based on alignment with the concatenated sequences of 37 baculovirus core genes, we found that the closest relative to PeluSNPV was Clanis bilineata nucleopolyhedrovirus, another sphingid-infecting alphabaculovirus. An interesting feature of this novel genome was the presence of a putative nucleotide metabolism enzyme-encoding gene (*pelu112*). The *pelu112* gene was predicted to be a fusion of *thymidylate kinase* (*tmk*) and *deoxyuridine triphosphatase* (*dut*), and this fused genes appears to have also been acquired convergently by two other distantly related baculoviruses. Moreover, phylogenetic analysis indicated that baculoviruses have independently acquired *tmk* and *dut* several times during their evolution from different sources. In order to test whether the expression of a *tmk-dut* fusion gene by a baculovirus that naturally lacks it would result in an adaptive gain, we inserted two homologs of the *tmk-dut* fusion gene into the Autographa californica multiple nucleopolyhedrovirus (AcMNPV) genome. The recombinant baculoviruses produced viral DNA, virus progeny, and some viral proteins earlier during *in vitro* infection and the yields of viral occlusion bodies were increased 2.5-fold when compared to the parental virus. Interestingly, both enzymes appear to

retain their active sites, based on separate modeling using previously solved crystals tructures. We therefore suggest that the retention of these *tmk-dut* fusion genes by certain baculoviruses could be related to accelerating virus replication. The hypothetical mechanism is likely related to synchronizing the cell cycle state, controlling the cellular nucleotide pool size (dUTP/dTTP ratio), or altering the expression or function of cellular nucleotide metabolism enzymes.

Keywords: Baculovirus, PeluSNPV, AcMNPV, *thymidylate kinase* (*tmk*), *deoxyuridine triphosphatase* (*dut*), horizontal gene transfer.

Este capítulo ainda não foi publicado. **Genome sequence of Perigonia lusca single nucleopolyhedrovirus (PeluSNPV): insights on the evolution of a nucleotide metabolism enzyme in the family *Baculoviridae*.** Daniel M. P. Ardisson-Araújo, Rayane Nunes Lima, Fernando L. Melo, Rollie Clem, Ning Huang, Sônia Nair Báo, Daniel R. Sosa-Gómez, Bergmann M. Ribeiro.

2. Introduction

Large double-stranded DNA viruses exhibit high genomic plasticity and primarily evolve by both horizontal gene transfer (HGT) and gene duplication/loss (Becker, 2000; Monier *et al.*, 2007). In many cases, viruses take advantage of an existing cellular pathway and fully or partially incorporate it into their genome (Monier *et al.*, 2007). With the increasing availability of genome sequence data, HGT events have been extensively documented in several viral families. This is particularly true for members of *Baculoviridae*, a family of dsDNA viruses infective mostly to larval stages of lepidoptera (moths and butterflies) (Jehle *et al.*, 2006).

More than 500 different types of genes have been found in the genomes of the 70-plus baculoviruses from different species that have been sequenced to date (Miele *et al.*, 2011), and many of them seem to be products of HGTs (Katsuma *et al.*, 2008). Exactly how most of these genes have been fixed in the genomes of baculoviruses still remains unclear (Rohrmann, 2013). For instance, an interesting but poorly studied group of genes acquired by baculoviruses are those related to nucleotide metabolism. Various baculoviruses contain homologs to *dUTP diphosphatase* (*dut*), *ribonucleotide-diphosphate reductase* (*rnr*), and *thymidine monophosphate kinase* (*tmk*), but none of these have been characterized at the molecular level and there is no evidence of fitness changes associated with them. Moreover, it has been suggested that baculoviruses have independently acquired *dut* and *rnr* genes more than once during their evolution (Herniou *et al.*, 2003).

Several viruses including baculoviruses, asfarvirus, herpesviruses, poxviruses, and certain retroviruses encode deoxyuridine triphosphatase (dUTPase) and/or thymidine monophosphate kinase (TMK) enzymes in their genome. However, it is unclear why these viruses encode an enzyme that is already encoded by the host cell. The enzyme dUTPase is conserved in prokaryotic and eukaryotic cells and such conservation is thought to be related to the shared inability of DNA polymerases in discriminating between dUTP and dTTP during DNA synthesis (Dube *et al.*, 1979). The enzyme TMK participates in both the *de novo* and the salvage dTTP biosynthesis pathways(Reichard, 1988). The misincorporation of dUTP in lieu of dTTP can lead to either deleterious mutations in the cell genome or to futile repair cycles and DNA breakage events that kill the cell (Ladner, 2001). Therefore, dUTPase activity associated with dTTP

biosynthesis pathway enzymes (*e.g.*TMK) are an essential preventive DNA repair mechanism that hydrolyses dUTP to dUMP and PPi and thereby plays a role in both lowering the dUTP/dTTP ratio and in providing substrate for the major biosynthesis pathway of dTTP (Mustafi *et al.*, 2003). Other roles for dUTPases have been demonstrated including transposase-like activity, regulation of the immune system, autoimmunity, and apoptosis, suggesting that they also perform regulatory functions (Penades *et al.*, 2013).

The baculovirus Perigonia lusca single nucleopolyhedrovirus (PeluSNPV) is a natural pathogen that was previously discovered infecting the half-blind sphinx moth *Perigonia lusca ilus* (Lepidoptera: Sphingidae) in 1988 (Sosa-Gómez *et al.*, 1994). So far, *P. lusca* does not present great agricultural interest, despite causing occasional damage on crops of Paraguay tea (*Ilex paraguariensis*) and Krug's holly (*I. krugiana*), genipapo (*Genipaamericana*), and coffee (*Coffeaarabica*) in Brazil (Primo *et al.*, 2013), Argentina, Puerto Rico, Cuba, and USA (The Natural History Museum, http://www.nhm.ac.uk). In previous work, the half-blind sphinx-infecting baculovirus was structurally described (Sosa-Gómez *et al.*, 1994); however, neither genomic organization nor phylogenetic relationships of the virus have been described. In this work, we sequenced the complete genome of PeluSNPV and established its phylogeny to other baculoviruses. Furthermore, a *tmk-dut* fused gene was found in the PeluSNPV genome which led us to the reconstruction of the phylogenetic history of *dut* genes in the *Baculoviridae*. When both the PeluSNPV *tmk-dut* fused gene and another baculovirus homolog were inserted into the baculovirus Autographa californica multiple nucleopolyhedrovirus (AcMNPV), which naturally lacks a *dut* gene, accelerated virus progeny production, virus genome replication, and viral gene expression were observed. These results lead us to hypothesize that the reason why

nucleotide metabolism genes, especially *tmk-dut*, are fixed in some baculovirus genomes may be due their ability to control the size of the cellular nucleotide pool, enabling faster virus replication.

3. Results

3.1. Structural analysis, genome features, and phylogeny of PeluSNPV.

For structural analysis, we performed a scanning electron microscopy (SEM) of purified occlusion bodies (OBs) of PeluSNPV. Mature OBs with non-regular shape and size were observed (FIG.1A). Immature OBs revealed singly enveloped nucleocapsid occlusion spaces (inset, Fig. 1A) as previously described (Sosa-Gómez *et al.*, 1994). Furthermore, restriction analysis of the virus DNA revealed that PeluSNPV was probably a novel virus since no similar restriction profile was found in the literature (Fig. 1B). Distinctions among species of the *Baculoviridae* have been based on DNA restriction endonuclease fragment patterns and comparisons of nucleotide and predicted amino acid sequences from various genes. A proposed species demarcation criterion was published in 2006 that is based on pairwise nucleotide distances estimated using the Kimura 2-parameter model of nucleotide substitution (Jehle *et al.*, 2006).

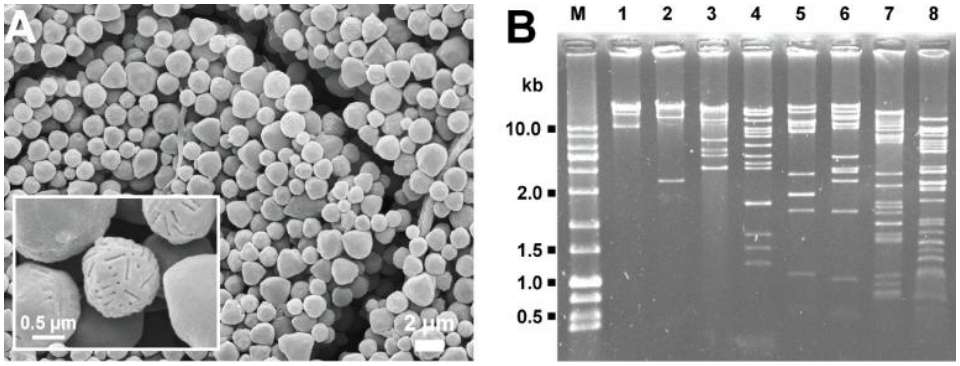

Figure 1. Structural analyses of PeluSNPV. (A) Scanning electron microscopy of purified polyhedral occlusion bodies (OBs) with non-regular shape and size. Immature OBs are inset. Moreover, singly embedded rod-shaped nucleocapsid spaces are shown. (B) Agarose gel electrophoresis-resolved DNA fragments digested with each *ApaI (lane 1), BamHI (lane 2), PstI (lane 3), XbaI (lane 4), XhoI (lane 5), BglII (lane 6), NsiI (lane 7), or ClaI (lane 8).* Molecular weight marker (lane M).

The entire genome of PeluSNPV was sequenced using 454 technology (Genbank accession number KM596836). Over 18,807 single-end reads were obtained. After size and quality trimming, 18,355 reads (mean size of 356.6 ±147.1 bp) were used for *de novo* assembly with a pairwise identity of 96.3 %. The mean coverage was 50.4±12.5 bases/site. The PeluSNPV genome was shown to contain 132,831 bp with a G+C content of 39.6 %. We found 145 putative ORFs encoding polypeptides with at least 50 amino acid residues (Table S1). Eighteen of these were shown to be unique in baculoviruses with no predicted motifs (*pelu004*, *pelu006*, *pelu010*, *pelu017*, *pelu018*, *pelu026*, *pelu035*, *pelu048*, *pelu054*, *pelu055*, *pelu089*, *pelu099*, *pelu100*, *pelu101*, *pelu119*, *pelu120*, *pelu140*, and *pelu144*) and only two homologous regions (hrs) with approximately 1,000 bp each were observed. All of the currently defined 37 baculovirus core genes were found and, based on phylogenetic analysis using the concatenated

alignment of the core genes from the completely sequenced baculoviruses (Table S2), PeluSNPV was found to belong to the genus *Alphabaculovirus* and clustered with Clanis bilineata nucleopolyhedrovirus (ClbiNPV), the first group II sphingid-infecting alphabaculovirus sequenced (Fig. 2). The nucleotide identity of PeluSNPV core genes (*i.e.* the 37 genes) with the closest relative ClbiNPV was 58%. Branch length separating this virus from its closest relatives is in a range that is comparable to the branch lengths separating viruses in other recognized alphabaculovirus species. Furthermore, many inversions, deletions, and insertions were observed in the genome of these closely related species when the gene content of PeluSNPV was compared to both ClbiNPV (Fig. 3 A) and AcMNPV (Fig. 3B) by gene parity plot. The gene order was not strictly conserved between PeluSNPV and ClbiNPV and four major inversions were detected (Fig. 3A). Although these sphingid-isolated viruses are closely related to each other, each contains several unique genes. The pairwise distances of the viral sequences of PeluSNPV to other alphabaculoviruses for both single locus and concatenated alignment are well in excess of 0.05 substitutions/site fulfilling all the criteria for a novel baculovirus species.

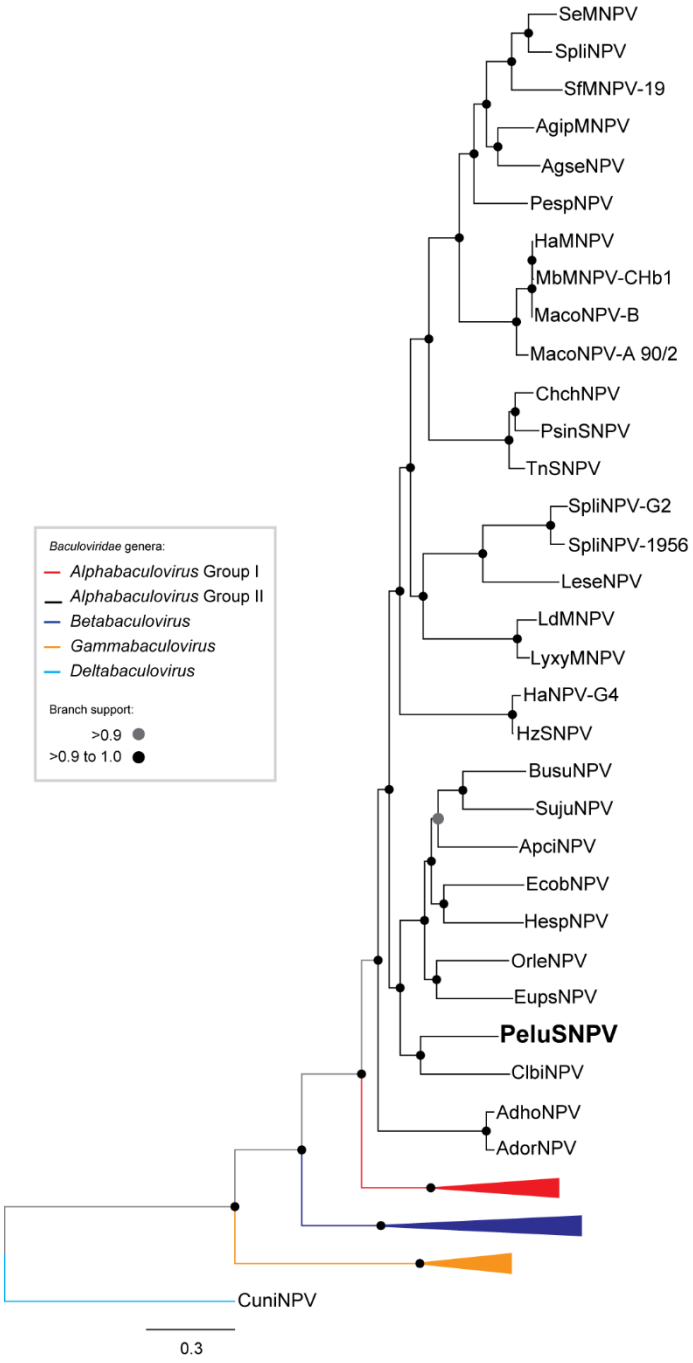

Figure 2. PeluSNPV is a Group II alphabaculovirus. Maximum likelihood inference based on the concatenated amino acid sequences of 37 core proteins of all complete baculovirus genomes (Table S2). The branch support was determined by a SH-like method. Some branches were collapsed for clarity: *Gammabaculovirus* (orange), *Betabaculovirus* (dark blue), and group I *Alphabaculovirus* (red). The deltabaculovirus

CuniNPVwas used as the root (light blue). PeluSNPV (boldface) belongs to the genus *Alphabaculovirus* and clustered with another sphingid-infecting group II alphabaculovirus, ClbiNPV.

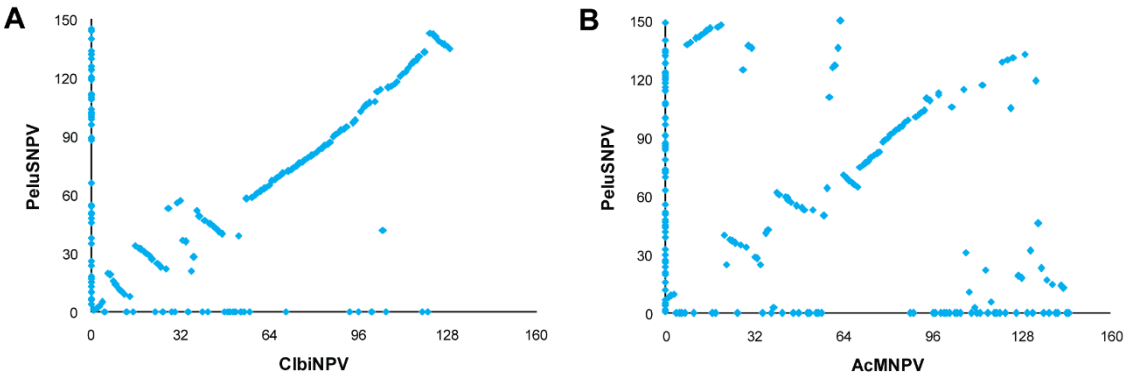

Figure 3. Gene content and sinteny of PeluSNPV compared to other two species. (A) PeluSNPV was compared to ClbiNPV, another sphingid-infecting baculovirus. (B) PeluSNPVwas compared to the baculovirus type species, AcMNPV.

3.2. Gene content

Several known examples of auxiliary genes were observed in the PeluSNPV genome. For instance, both *cathepsin* and *chitinase* were found in the genome in an opposite orientation, as commonly found in other baculovirus genomes. The putative chitinase presents a KTEL motif at the very end of the C-terminal region, which is related to retention into the ER. The presence of these genes is consistent with the post-mortem phenotype observed for the host caterpillar infected with PeluSNPV, which includes both body melanization and liquefaction of internal tissues (data not shown). The *iap-2* (*pelu064*) and *iap-3* (*pelu102*) genes, which are usually present in the genomes of group II alphabaculoviruses and are involved in the anti-apoptotic response induced by virus

infection, were also observed. However, the predicted *iap-3* (*pelu102*) homolog lacks one of the two commonly conserved Baculovirus IAP Repeat (BIR) domains at the N-terminal region (data not shown), which is involved in protein-protein interactions (Hinds *et al.*, 1999). Furthermore, we found a homolog of non-structural (NS) densovirus gene, *pelu104*. Homologs of this gene were previously found in three betabaculovirus genomes including Choristoneura occidentalis granulovirus (*choc025*) (Escasa *et al.*, 2006), Cryptophlebia leucotreta granulovirus (*crle009*) (Lange & Jehle, 2003), and Erinnyis ello granulovirus (*erel057* and *erel100*) (Ardisson-Araujo *et al.*, 2014a). To our knowledge, PeluSNPV is the first alphabaculovirus harboring a densovirus-related gene. The phylogenetic reconstruction revealed that PeluSNPV probably acquired it from a betabaculovirus (data not shown). The fitness effects of this gene are unknown, but a *Helicoverpa armigera*-associated densovirus was found to protect the host insect from both baculovirus and *Bacillus thuringiensis* infection (Xu *et al.*, 2014). Moreover, a homolog of *he65* (RNA ligase-like gene) was also found in the PeluSNPV genome, *pelu124*. In a previous study, we reconstructed the phylogenetic history of *he65* and found that it is present in several baculovirus and two entomopoxvirus genomes. Importantly, a large and recurrent deletion observed at the C-terminal region of the putative baculovirus proteins has also been observed in the putative Pelu124 (Ardisson-Araujo *et al.*, 2014a). The phylogenetic analyses clustered *pelu124* with both group II alphabaculovirus and entomopoxvirus genes, while the closest baculovirus relative of PeluSNPV (*i.e.* ClbiNPV) lacks *he65* ortholog.

3.3. Genes related to nucleotide metabolism

Genes encoding both the large and small subunits of ribonucleotide reductase (RNR) were found in the PeluSNPV genome, *pelu145* and *pelu126*, respectively. Ribonucleotide reductase catalyzes the rate-limiting step for deoxyribonucleotide production required for DNA synthesis. The enzyme is a tetramer consisting of two large and two small subunits (Huang & Elledge, 1997). Several baculoviruses and other arthropod-related viruses contain these genes in their genomes including the white spot syndrome virus (van Hulten *et al.*, 2001). The presence of these genes has been also associated with the presence of *dut* genes in baculovirus genomes (Herniou *et al.*, 2003) but some *dut*-harboring betabaculoviruses lack the RNR enzyme (*e.g.* ErelGV) (Ardisson-Araujo *et al.*, 2014a).

The putative ORF *pelu112* was found to be a nucleotide metabolism gene with some peculiar features. Firstly, *pelu112* was found to be a fusion of two putative genes. The predicted N-terminal region was related to the *cypo016* gene of the baculovirus Cydia pomonella granulovirus (CypoGV), which has identity with a *thymidylate kinase* (*tmk*, Fig. 4A) whereas the predicted C-terminus was related to *dut* (Fig. 4B). Several secondary structures were conserved when both regions were compared to previously solved crystal proteins. Moreover, *tmk* and *dut* homologs are present in many other baculovirus genomes as separated ORFs or, in the case of the latter one, often fused to other genes. Secondly, *pelu112* has homologs in two other distantly related baculoviruses, ErelGV (*erel005*) (Ardisson-Araujo *et al.*, 2014a) and Orgyia pseudotsugata multiple nucleopolyhedrovirus (OpMNPV) (*op031*) (Ahrens *et al.*, 1997)

(Fig. 4) with pairwise identity of , 90.2% and 74.1% respectively. The identities were obtained by MAFFT alignment.

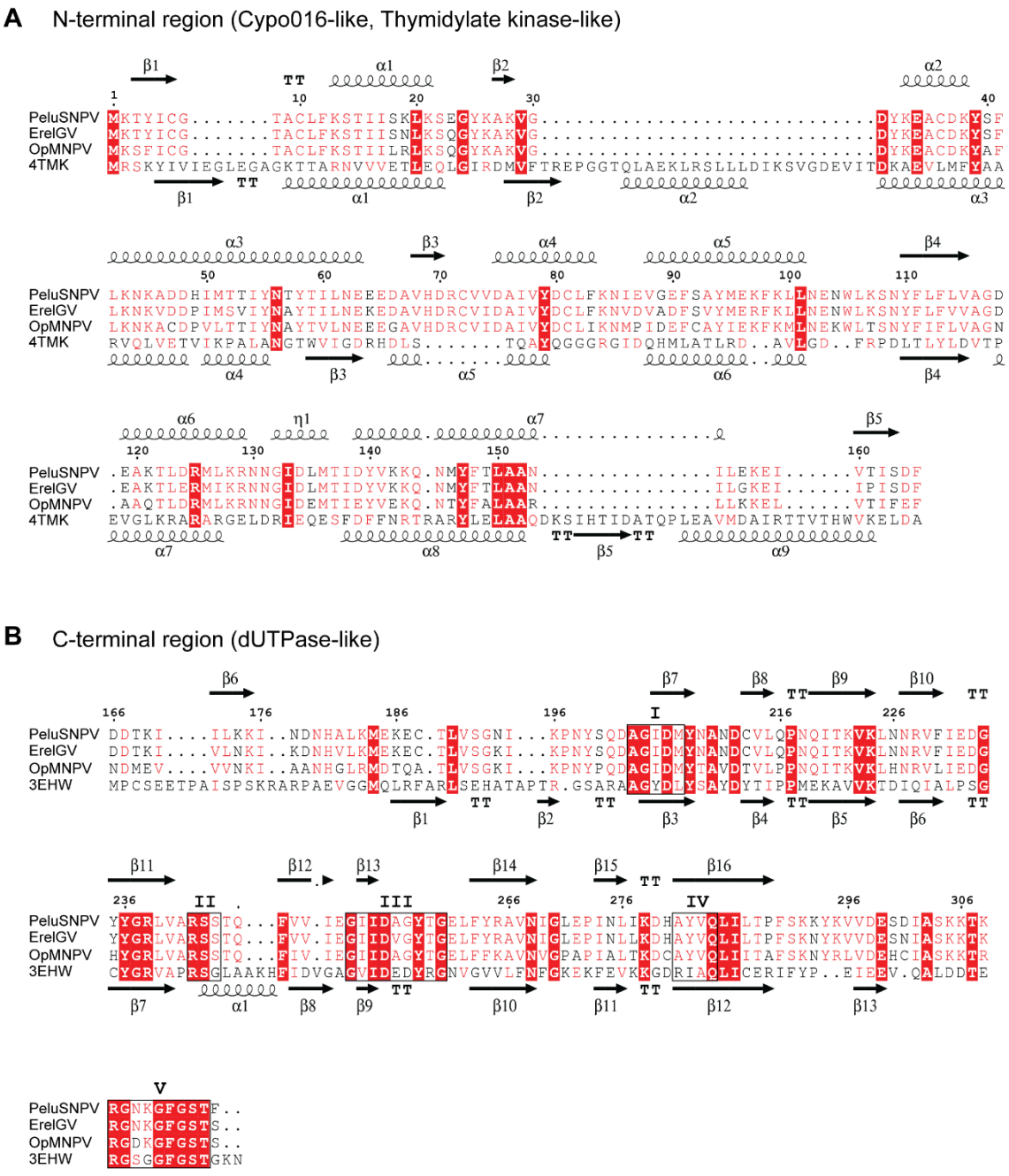

Figure 4. Individual alignments of both TMK and dUTPase regions of PeluSNPV, ErelGV, and OpMNPV against proteins with crystal solved structures. (A) Predicted N-terminal region presents homology to Cypo016, a putative thymidylate kinase enzyme.

(B) Predicted C-terminal region presents homology to trimeric dUTPases. The conserved motifs are boxed in black lines from I to V. The predicted secondary structures are shown for both Pelu112 regions and the proteins with crystal solved structures. α/spirals: α-helices; β/arrows: β-sheet; tt: turns; dashed lines: no secondary structure found; red box: strictly conserved residues.

3.4. Phylogenetic analysis of *pelu112* gene

We performed separate phylogenetic reconstructions of both regions (*tmk* and *dut*) of *pelu112* (Fig. 5). In the *tmk* dataset, we included genes related to entomopoxvirus, nudivirus, and to the mealworm disease-associated apicomplexan *Gregarina niphandrodes* obtained by BLASTX. The ErelGV-, OpMNPV- and PeluSNPV-derived genes clustered together, suggesting a common ancestry (Fig. 5A). The closest relatives were both nudivirus and apicomplexan genes. Betabaculovirus-derived *tmk* genes (except ErelGV) clustered together and the same occurred with alphabaculovirus group II genes. The unique exception for alphabaculoviruses was the ClbiNPV gene, suggesting an independent HGT event. We confirmed this by looking at the gene context in the genome and as expected, all HGT events presented different genomic contexts (data not shown).

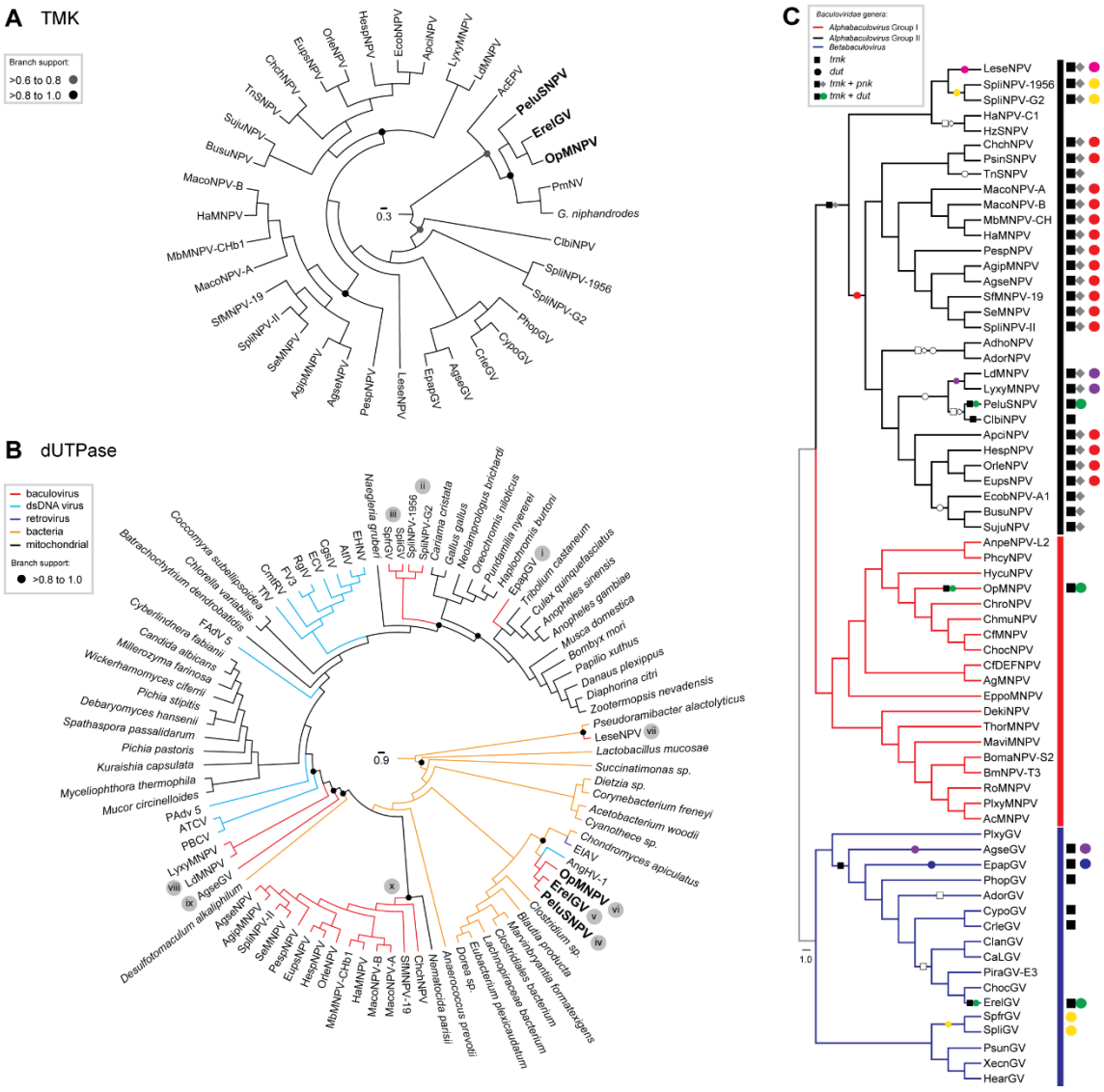

Figure 5. Phylogeny and evolution of both TMK and dUTPase regions in the family *Baculoviridae*. (A) Phylogeny of cypo16-like, the N-terminal portion of *tmk-dut* fused gene. ErelGV, OpMNPV, and PeluSNPV-derived proteins clustered together, indicating common ancestry. (B) Phylogeny of dUTPases in the family *Baculoviridae*. Several dUTPases clustered and seemed to be shared by several group II alphabaculoviruses. The putative independent acquisitions are numbered from *i* to *x*. (C) Based on the hypothetical phylogeny trees, the history of gain and loss of both *tmk* and *dut* in the family *Baculoviridae* were described. For this phylogenetic analysis, we used the concatenated alignment of 37 core genes of alpha and betabaculoviruses. Filled and

empty symbols represent gain and loss events, respectively. Similar events of *dut* acquisitions (circles) are shown with the same color. All the trees were midpoint rooted and presented as cladogram for clarity.

We carried out a similar phylogenetic analysis using the predicted protein sequence of several *dut* genes from bacteria, viruses, and mitochondrial isoform genes. We found that many group II alphabaculovirus *dut* genes clustered together forming a well-supported monophyletic clade with a fungus mitochondrial gene being the likely ancestor (Fig. 5B). Conversely, some baculovirus genes were found to be spread along the tree depicting at least nine predicted HGT events from several sources including other baculoviruses (Fig 5B and C). The *dut* gene of Epinotia aporema granulovirus (EpapGV) seemed to be acquired from an insect mitochondrial isoform gene (*i*). The *dut* genes of Spodoptera litura granulovirus (SpliGV), Spodoptera frugiperda granulovirus (SpfrGV), Spodoptera litura nucleopolyhedrovirus AN1956 (SpliNPV-1956), and Spodoptera littoralis nucleopolyhedrovirus II (SpliNPV-II) clustered together and it seems to be product of a double HGT event (*ii* and *iii*). Firstly, the gene was probably acquired from an amoeba-related mitochondrial isoform by the ancestor of either SpliGV and SpfrGV or SpliNPV-1956 and SpliNPV-II. The second event may have occurred during a co-infection scenario of a *Spodoptera sp*. host by both ancestors. Three other independent acquisitions (*iv*, *v*, and *vi*) seemed to take place in PeluSNPV, ErelGV, and OpMNPV evolution, that formed a dissimilar well-supported subclade closely related to bacteria-, lentivirus-, and adenovirus-derived *dut* genes (Fig. 5B). This acquisition happened probably once in the ancestor of one of those species (*i.e.* PeluSNPV, OpMNPV, ErelGV) and was transferred to the other baculoviruses during co-infection events. For instance, both PeluSNPV and ErelGV are sphingid-infecting

baculovirus and their ancestors could potentially infect the same host. Another event appears to have occurred in Leucania separata nucleopolyhedrovirus (LeseNPV) (*vii*), with its closest relative being a bacterium. Finally, Lymantria xylina multiple nucleopolyhedrovirus (LyxyMNPV), Lymantria dispar multiple nucleopolyhedrovirus (LdMNPV) (*viii*) and Agrotis segetum granulovirus (AgseGV) (*ix*) appear to have independently acquired their homologs from unknown ancestors. To further substantiate our findings, we examined the genomic context of the baculovirus*dut* genes, since unrelated HGT usually occurs at different genomic loci. As expected, all HGT events presented different genomic contexts (data not shown).

The *tmk* genes are found in three different manners in the baculovirus genomes: fused to either a *polynucleotide kinase 3'-phosphatase* (*pnk*, previously annotated as a *nicotinamideriboside kinase* 1, nrk-*1*) or *dut*, or alone (Fig. 5C). In group II alphabaculoviruses, the gene is usually fused to the N-terminal portion of *pnk* (closed square/diamond, Fig. 5C). The unique exception was in ClbiNPV, where no *pnk* is found and the genomic context is different when compared to the other viruses (data not shown). Therefore, we concluded that some species lost the *tmk* gene during evolution (open square/diamond, Fig. 5C) and reacquired it independently from an undisclosed source (*e.g.* ClbiNPV and PeluSNPV) (Fig. 5C). On the other hand, only in PeluSNPV, ErelGV, and OpMNPV was a *tmk* gene found fused to the N-terminal region of a *dut* gene (square/circle, Fig. 5C). Finally, *tmk* was found with no fusion in most betabaculoviruses (single square, Fig. 5C).

3.5. Two *tmk-dut* genes were expressed and localized distinctly in infected cells

We engineered the type baculovirus, Autographa californica multiple nucleopolyhedrovirus (AcMNPV), by inserting separately either *pelu112* or *erel005* with an N-terminal HA tag (Fig. 6A). AcMNPV naturally lacks *dut*, *tmk*, and any other nucleotide metabolism genes. The genes were inserted under the transcriptional control of a constitutive insect promoter (*Drosophila melanogaster heat shock protein 70* gene promoter) (Ardisson-Araujo *et al.*, 2015a). Immunoblotting analysis confirmed that both *pelu112* and *erel005* were expressed as fusions and not as cleaved proteins, based on their migration. Although both proteins have similar predicted molecular masses (37.5 kDa), *pelu112* produced a product that migrated more slowly compared to *erel005* (Fig. 6B). Time course analysis of the recombinant virus infections revealed that the proteins were first detected at 12 h p.i. and accumulated during infection progression (Fig. 6C). As a loading control, an over-exposure-derived unspecific reactive band is shown. By confocal microscopy at 24 h p.i., Pelu112 was found close to the plasma membrane and present in the cytoplasm, and the nucleus ring-zone, while Erel005 was mostly near the plasma membrane and in the cell cytoplasm (Fig. 6D).

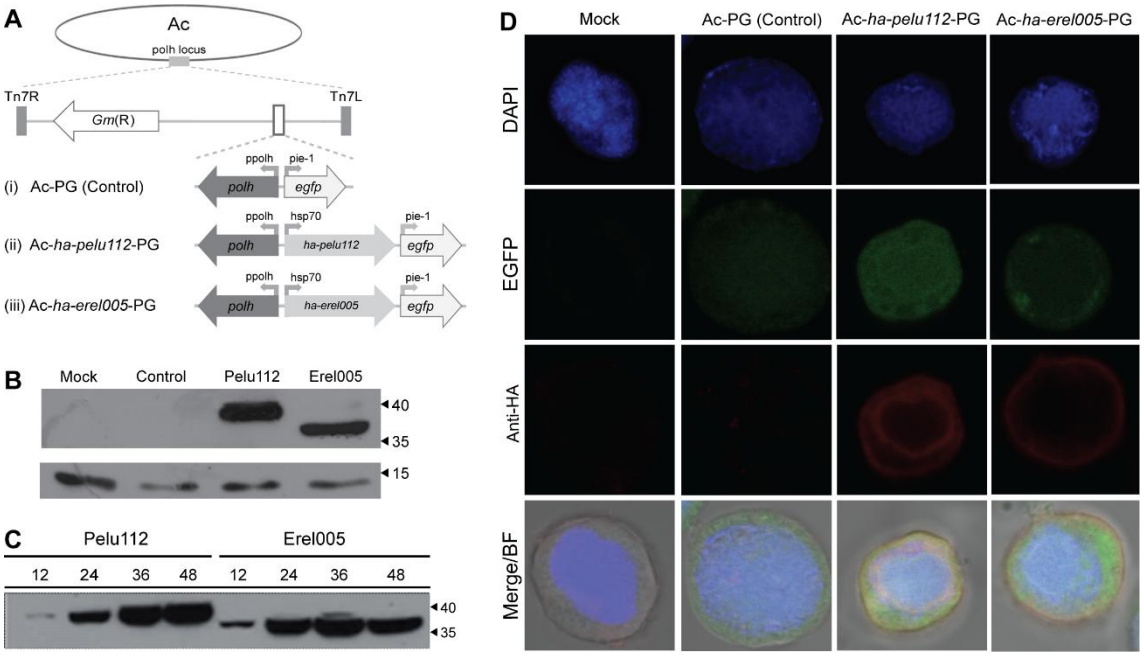

Figure 6. Schematic representation of engineered recombinant viruses, expression of HA-Pelu112 and HA-Erel005 proteins, and cytolocalization analyses. (A) The HA-tagged genes were inserted into the AcMNPV genome under the control of an insect constitutive promoter (hsp70). (B) Cells were mock-infected or infected with (i) Ac-PG (Control), (ii) Ac-*ha-pelu112*-PG (Pelu), or (iii) Ac-*ha-erel005*-PG (Erel) at an MOI of 0.01. Cells were harvested at 48 h p.i., and the total proteins were analyzed by immunoblotting with anti-HA antibody. An over-exposure-derived unspecific reactive band is shown as a loading control. (C) Expression kinetics of HA-tagged proteins were assessed by immunoblotting. (D) Cytolocalization in virus-infected Sf9 cells. Images of virus-infected cells (MOI of 10) were photographed at 24 h p.i. using confocal laser scanning microscopy. Image panels show the red (anti-HA secondary antibody), green (GFP expressed by all recombinant viruses), and blue (DAPI) fluorescent channels. Overlays of all channels and the bright-field images are also shown (MERGE/BF).

3.6. *tmk-dut* expression accelerated AcMNPV progeny production

In order to check whether expression of *pelu112* or *erel005* could influence baculovirus infection, we looked at *tmk-dut*-expressing virus progeny production *in vitro* using Sf9 cells. Interestingly, the recombinants expressing either *pelu112* or *erel005* produced higher levels of BV at 24 and 48 h p.i. than the control virus, although the final titers were similar at 72 and 96 h p.i. (Fig. 7A). For *pelu112*-expressing virus, the increase was 8.6- and 10.4-fold higher at 24 and 48 h p.i. respectively when compared to the parental virus, while for the *erel005*-expressing virus, the increase was 6.8- and 7.4-fold at the same times. Moreover, the yields of occlusion bodies (OB) were increased 2.5-fold in the *tmk-dut*-fused-expressing viruses compared to the control (Fig. 7B). It is important to note that in this experiment only OB production was monitored, not the ability to occlude virions.

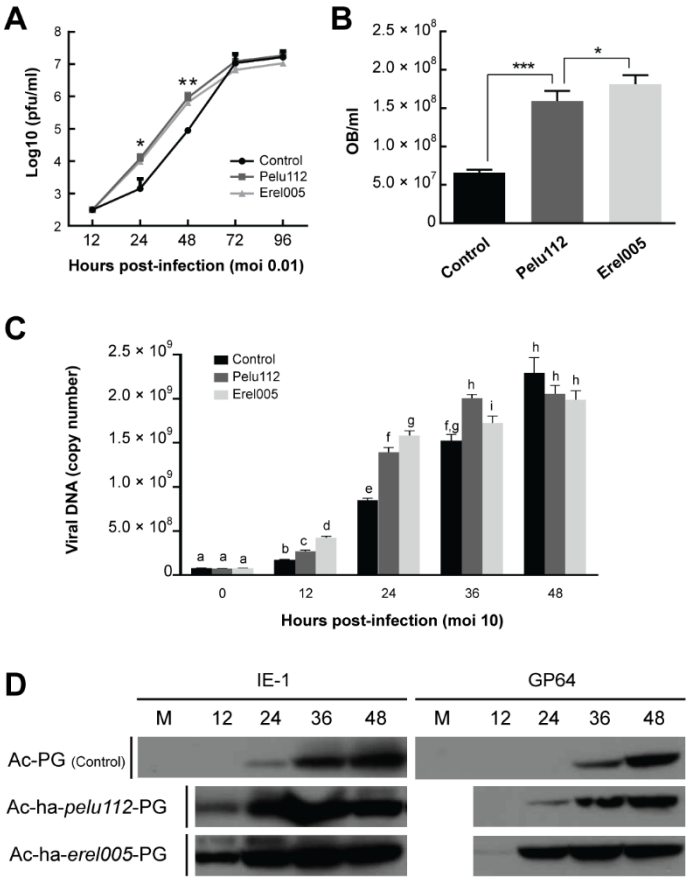

Figure 7. Expression of HA-Pelu112 or HA-Erel005 accelerated AcMNPV replication, viral DNA synthesis, and viral protein expression. (A) Analysis of BV production by endpoint dilution assays. Titers were determined from supernatants of cells infected with parental Ac-PG (Control), Ac-*ha-pelu112*-PG (Pelu), or Ac-*ha-erel005*-PG (Erel) (MOI of 0.01) at the designated time points in triplicate. Statistical differences at 24 and 48 h p.i. obtained by unpaired T-test are shown (*p* values: *, $p \leq 0.01$; **, $p \leq 0.001$). (B) Yields of occlusion bodies (OB) were increased 2.5-fold in the recombinant viruses. OBs were purified from Sf9 cells infected with the respective viruses (MOI of 5) at 120 hp.i.. Bar heights indicate the averages of four repeats, and the error bars represent the standard deviations. Statistical differences by unpaired T-test are shown (*p* values: ***, $p \leq 0.0001$; *, $p \leq 0.01$). (C) Cells were infected (MOI of 10) with the indicated viruses and at 0, 12, 24, 36, and 48 h p.i. total intracellular DNA was purified and analyzed by

real-time PCR in three repeats. Statistical difference by unpaired T-test are shown by letters above the bar heights. Different letters indicate that statistical difference exists. (D) The fused genes accelerated both IE-1 and GP64 expression during *in vitro* virus infection when compared to the control virus. Lysates obtained from the same number of cells was loaded in each lane. Cells were infected with the indicated viruses (MOI of 0.01) and at 0, 12, 24, 36, and 48 hpi total cellular proteins were analyzed by immunoblotting with specific anti-IE-1 or anti-GP64 antibodies.

3.7. AcMNPV replication and IE1 and GP64 expression were accelerated by the *tmk-dut* genes

Since homologs of *pelu112* and *erel005* are hypothetically thought to play roles in nucleotide biosynthesis pathways, we examined viral DNA replication during recombinant infection. Viral DNA replication was accelerated during recombinant infection *in vitro* and remained higher through 36 h p.i. (Fig. 7C). At 12 and 24 h p.i., the *erel005*-expressing virus produced more viral DNA than either the *pelu112*-expressing virus or the parental virus. However, at 36 h p.i. the recombinant harboring *pelu112* accumulated more DNA than the two others, while the *erel005*-expressing virus remained higher than the control. By 48 h p.i., there was no significant difference in the levels of viral DNA produced by any of the viruses. We also examined the levels of two essential virus proteins, IE-1 (the major alphabaculovirus transcription factor) and GP64 (the envelope fusion protein). Both proteins were detected earlier in cells infected with the *tmk-dut*-fusion-expressing viruses than with the control virus, consistent with the results observed for viral DNA replication and BV production (Fig. 7D).

3.8. Homology modeling

In order to determine whether *pelu112* and its homologs (*op031* and *erel005*) potentially encode functional proteins and this activity could be related to the viral performance infection change, we performed an alignment against homologs using solved crystal structures (Fig. 4). Both TMK (Fig 4A) and dUTPase (Fig. 4B) presented all the amino acid residues responsible for the enzymatic activity despite of presenting few variations. We also built a 3D model of each domain for the predicted amino acid sequence of Pelu112. The identity between the target sequences (N- and C terminal regions) and their templates were 27.15 % (PDB ID: 4TMK) and 28.06% (PDB ID: 3EHW), respectively. The Ramachandran plot of TMK region showed 92% residues in favored region, 5.52% in allowed region and 2.45% outliers (Fig. S1B). Whereas the dUPTase region showed 92% residues in favored region, 6% in allowed region, and 2% outliers (Fig. S1B). The overall structure of both TMK and dUTPase homology models were similar to that of the templates. The TMK-like enzyme at the Pelu112 N-region (FIG. 8A) has an α/β fold with a three-stranded parallel β-sheet surrounded by seven α-helices, similarly to other TMKs (Yan & Tsai, 1999). On the other hand, the Pelu112 C-terminal core is a putative homotrimer composed of β-strands (12 strands) (Fig. 8B and Fig. S1C). The dUTPase had a sequence homology to trimeric dUTPases and presented all the five conserved motifs commonly found intrimeric dUTPases (Fig. B). Moreover, the N-terminal region of the monomer is projected outward leaving it free to be fused to other proteins such as TMK (data not shown). A fusion model is also proposed (Fig. 8D and Fig. S1D).

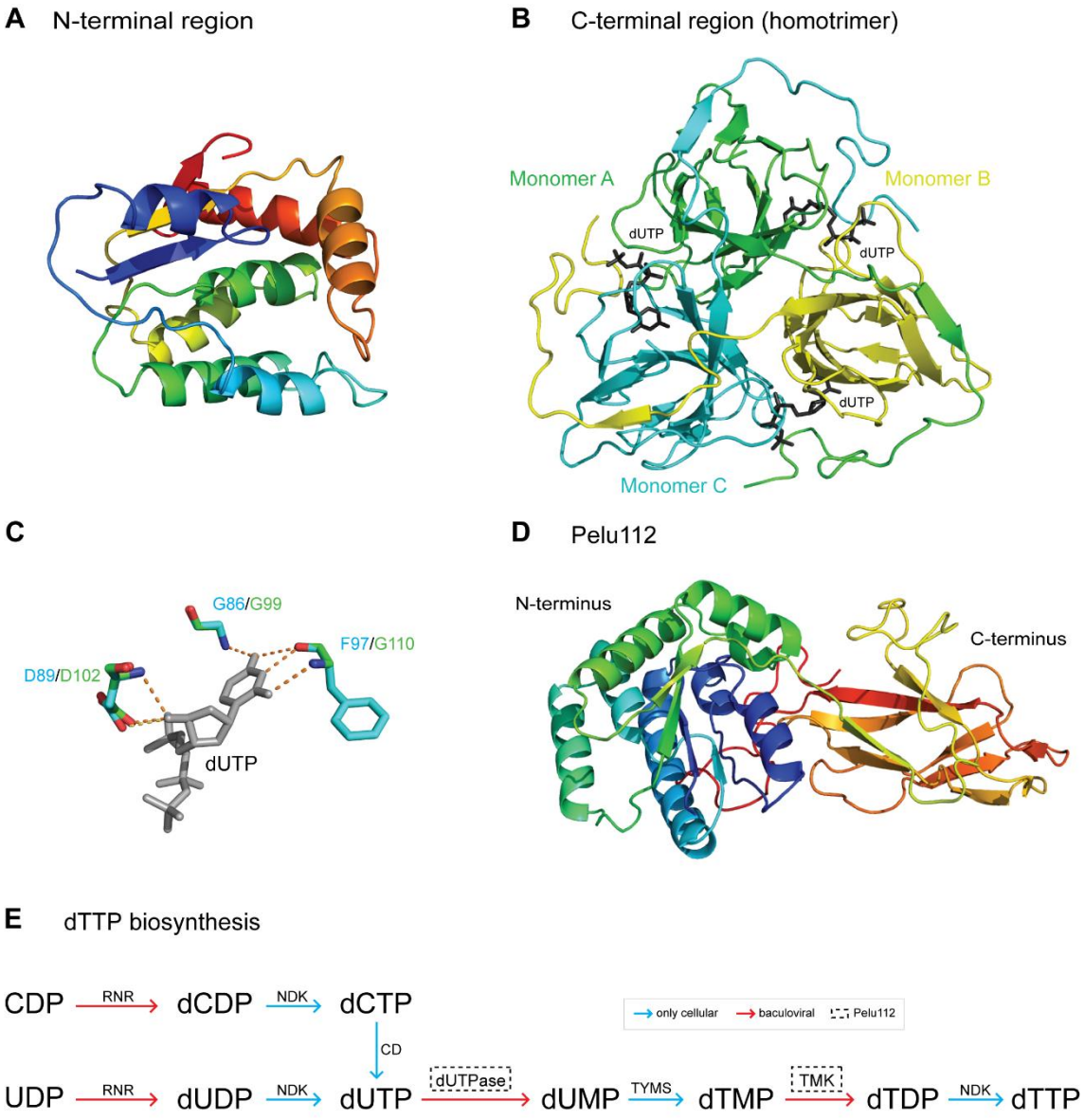

Figure 8. Homology modeling of Pelu112. (A) N-terminal region presents homology to thymidylate kinase. The model obtained presents several α-helixes as commonly is found in this enzyme. (B) Homotrimer model proposed for the C-terminal region of Pelu112. The three monomers interacting with their substrates (dUTP in black) are shown. (C) Conserved catalytic site of the modeled dUTPase interacting with dUTP (dashed lines). The template crystal used for the proposed model is shown in green overlapping the proposed model in blue. Although we identified one amino acid substitution in Pelu112 (G110 to F97), the interacting region was clearly conserved and

remained stable through projecting the lateral chain to outside from the catalytic site. (D) Fused model TMK-dUTPase. Both the N-terminus and C-terminus are shown. All the proposed models were constructed using previously solved protein structures available in PDB database. (E) dTTP biosyinthesis pathway presention baculovirus-encoded enzymes highlighting the enzymes that are fused in the Pelu112 (dashed box) and their respective action on the path. RNR, ribonucleotide reductase; NDK, nucleoside diphosphate kinase; CD, cytosine deaminase; TYMS, thymidylate synthase, TMK, thymidylate kinase.

Based on this, we conclude that these motifs form a functional dUTPase active site and allow the C-terminal region of Pelu112 to form a trimeric quaternary structure with three active sites per trimer capable to interacting at the N-terminal region with other proteins (Fig. 8B). Moreover, we overlapped the catalytic site from both the template and the proposed model of the dUTPase (Fig. 8C, light green). Only one amino acid difference was observed in the catalytic site, a phenylalanine in Pelu112 rather than a glycine. Crucially, this amino acid substitution did not impact the interaction with dUTP due to the positioning of amino acid lateral chain. Therefore, it is reasonable to assume that *pelu112* encodes a *bona fide* TMK-dUTPase enzyme enzyme related to different steps of the dTTP biosynthesis pathway (Fig. 8E).

4. Discussion

The complete genome sequence of the *Perigonia lusca*-isolated group II alphabaculovirus PeluSNPV revealed that the virus is a new species most closely related to Clanis bilineata nucleopolyhedrovirus (ClbiNPV), another sphingid-infecting

virus. In the PeluSNPV genome, we found all of the 37 baculovirus core genes and many auxiliary genes including a densovirus-related non-structural homolog, *he65*-like, *chitinase*, *cathepsin*, *iap-2* and *iap-3*, and both small and large subunits of the *ribonucleotide reductase*. Moreover, the genome sequence revealed a peculiar nucleotide metabolism gene acquisition (*pelu112*) which was found to be a fusion of two other genes with separate homologs in other genomes, *thymidylate kinase* or *thymidine monophosphate kinase* (*tmk*) and *deoxyuridine triphosphatase* (*dut*), and this particular gene fusion seemed to be acquired independently by two other distantly related baculoviruses. Reconstructing the evolutionary history of both regions separately, we found that (*i*) this form of *tmk* seemed to be acquired several times during baculovirus evolution as a fusion or non-fused protein, while the *dut* has been acquired at least ten times. Furthermore, we have provided for the first time experimental evidences that expressing a fused nucleotide metabolism gene in a prototypic baculovirus that naturally lacks it resulted in accelerated *in vitro* virus progeny production, viral gene expression, and genome replication, as well as increased OB yields. Both enzymes retained tertiary structures predicted based on alignment with crystal-solved enzymes, which is strong, but not confirmed evidence of enzyme activity. Together, our results suggest that encoding a nucleotide metabolism gene homolog is beneficial for baculovirus replication and infection *in vitro*, and likely explains why these genes have been repeatedly acquired and retained during baculovirus evolution.

As a general rule, neither *tmk* or *dut* are essential for baculovirus infection given that several species lack them (Fig. 5C). However, the independent and recurrent acquisition of nucleotide metabolism genes, especially *dut*,from distinct taxonomic groups by baculoviruses and other viruses strongly suggests that there is a selective advantage for

viruses harboring these genes. Indeed, a gene that provides accelerated progeny production such as that observed for the recombinant viruses produced in this work would be probably fixed into the virus population along the course of evolution. Importantly, in this work we are not asking whether the enzyme activities are the main reason for the adaptive gain observed in the recombinant viruses, since we did not test for it. Even though we have shown that the fused enzymes retained their individual structures and catalytic site as well, our evolutionary question here is whether the presence and expression of a nucleotide metabolism gene into a prototype baculovirus that naturally lacks it may change the virus infection. Therefore, we have chosen this especial fused gene for two main reasons; firstly, the gene has being independently acquired three times during baculovirus evolution and secondly the gene is a fusion of two nucleotide metabolism genes.

There is no clear evidence indicating that this fusion would negatively impact the hypothetical enzyme activities. By homology modeling, we observed the possibility that fusion would have no allosteric impact on the chimeric structure and we did not observe cleavage products when the chimera was expressed during baculovirus infection. The fusion of nucleotide metabolism-related genes is also observed in the genome of the nimavirus White Spot Syndrome Virus (WSSV). The WSSV genome encodes a *thymidine kinase* (*tk*) fused to a *tmk* (Tsai *et al.*, 2000) despite only TK activity being demonstrated in the fused gene (Tzeng *et al.*, 2002).

In an attempt to understand and explain our results, we found in previously published work that the expression of cellular dUTPase is regulated by the cell cycle and is at higher levels in dividing cells than in non-dividing cells (Pardo & Gutierrez, 1990;

Strahler *et al.*, 1993). In the context of virus infection, uracil incorporation controlled by the expression of cellular dUTPase and enzymes related to dTTP biosynthesis could work as a weapon against viruses (Priet *et al.*, 2006). For baculoviruses, insect cells do not undergo synchronous division when cultivated *in vitro* as stable lineages (Braunagel *et al.*, 1998; Lynn & Hink, 1978) and hence the nucleotide pool size likely also varies between cells. Cultures with higher percentages of cells in middle and late S phase are more susceptible to baculovirus infection than cultures inoculated with virus in the G_2 phase (Lynn & Hink, 1978). An advantage for viruses to be able to replicate efficiently in a heterogeneous cell type tissue may allow them to establish infection more effectively in the host (Chen *et al.*, 2002; Steagall *et al.*, 1995).Therefore, a virus that harbors dUTPase, TMK, and other enzymes related to nucleotide metabolism (*e.g.* ribonucleotide reductase) could be able to better replicate in cells that are not in S phase by controlling the nucleotide pool size. In dividing cells, dUTPase activity would presumably not be necessary for the replication of several pathogens including herpersviruses, asfavirus, and several lentiviruses,while in non-dividing cells the virus replication is significantly reduced (Caradonna & Cheng, 1981; Lerner *et al.*, 1995; Oliveros *et al.*, 1999; Pyles *et al.*, 1992; Ross *et al.*, 1997; Threadgill *et al.*, 1993; Turelli *et al.*, 1996). On the other hand, the replication of dUTPase-minus lentivirus mutants was severely affected in non-dividing host cells (*e.g.* primary macrophages), with a decrease in virus load and an increase in viral DNA transition mutations (Turelli *et al.*, 1997). Interestingly, in the case of the four known betabaculoviruses that harbor nucleotide metabolism genes (*e.g. dut*, *tmk*, or *rnr*) each species possesses a *dut* gene that appears to have been captured on four independent occasions. Both AgseGV and EpapGV are known to present polyorganotropic pathology (Ferrelli *et al.*, 2012; Goldberg *et al.*, 2002) which means that the virus can spread throughout the insect body

and is not restricted to the midgut. Since cell division rates vary according to the tissue type, nucleotide metabolism genes could help viruses to overcome the non-dividing cell state of some tissues. Interestingly, EpapGV codes for a novel enzyme TMK (Ferrelli *et al.*, 2012) that seems to differ from the *tmk* gene addressed by this work. We did not find close relationship by BLAST search between them; therefore, the *tmk* gene found at the N-terminal region of Pelu112 has no clearly defined source. TMK enzymes can be found in several viruses from different families including *Asfaviridae*, *Herpesviridae*, and *Poxviridae* and some of them seem to be homologs. In vaccinia virus, the enzyme was not essential for virus replication and was able to complement the enzyme of a *Saccharomyces cerevisaetmk* mutant (Hughes *et al.*, 1991).

Along these lines, we found that HA-Pelu112 migrated more slowly in SDS-PAGE compared to HA-Erel005, despite both proteins having similar predicted molecular masses (37.55 kDa) and high pairwise amino acid identity (90.2 %). This difference in migration could be related to a type of post-translational modification such as phosphorylation. *Herpes simplex virus* 1 (HSV-1) dUTPase phosphorylation regulates viral virulence and genome integrity by compensating for the low cellular dUTPase activity in the central nervous system (Kato *et al.*, 2015). We found also that HA-Pelu112 and HA-Erel005 presented different patterns of cytolocalization upon infection progression. Pelu112 was observed in both cytoplasm and nucleus while the betabaculovirus-derived protein Erel005 was found only in the cell cytoplasm. Phosphorylation can lead to different patterns of cell localization (Nardozzi *et al.*, 2010). The dUTPase from Ophiusa disjungens nucleopolyhedrovirus was found to be in the cell nucleus at 24 h p.i., but at 72 h p.i. it was excluded from this compartment and diffusely scattered all over the cell (Lin *et al.*, 2012). It is relevant to note that

betabaculoviruses can cause nuclear disruption upon infection, making the infected cell a mixture of cytoplasm and nucleoplasm (Goldberg *et al.*, 2002; Lacey *et al.*, 2011).

There is no clear reason to explain why some viruses have a nucleotide metabolism gene and others lack it. The reason could be related to conditional expression or specificity of the host enzyme. For instance, *E. coli* dUTPase activity was not sufficiently active to exclude uracil from a dUTPase mutant bacteriophage T5 during infection, and about 3% of the thymine was replaced by uracil in viral progeny genomes (Warner *et al.*, 1979). In the case of HSV-1 with a *dut* gene deletion, the replication process was sufficiently complemented by a cellular dUTPase (Williams, 1988).

Overall, we cannot say for sure whether Pelu112 is an active enzyme but we have shown that both *tmk* and *dut* gene acquisition happened independently several times during baculovirus evolution, which also seems to be a convergent and common feature among other viruses (*e.g.* herpesvirus, iridovirus, phycodnavirus, adenovirus, and lentivirus). Moreover, we have shown that both regions of Pelu112 are structurally conserved and crucially, that the insertion of *tmk-dut* fused genes into the genome of AcMNPV, which does not normally express them, accelerated virus replication *in vitro*. We can only speculate that expression of *tmk-dut* accelerated replication by increasing the nucleotide pool size in non-dividing cells, making them a more permissive and less deleterious environment for virus replication. It would be interesting to study the function of nucleotide metabolism gene in its natural context by constructing a deletion virus and check for the enzyme activities in the fused protein and separately. However, our results have presented the first clues for explaining nucleotide metabolism gene fixation in baculovirus genomes. Overall, in this context, the identification and

sequencing of novel virus species or isolates, especially from countries with high diversity of flora and fauna such as Brazil, has provided a wider empirical database to help understand baculovirus evolution (Ardisson-Araujo *et al.*, 2014a; Ardisson-Araujo *et al.*, 2014b; Ardisson-Araujo *et al.*, 2015b; Craveiro *et al.*, 2015; Oliveira *et al.*, 2006).

5. Material and Methods

5.1. Virus purification

Insect cadavers of *P. lusca ilus* with symptoms of baculovirus infection were collected in mate tea crops from the South Brazil. The cadavers were kept in freezer and used for further OB purification (O'Reilly *et al.*, 1992).

5.2. Scanning electron microscopy (SEM) and genomic DNA restriction analyses

One hundred µl of the OB-containing suspension (10^9 OBs/ml of ddH_2O) were used for SEM according to previously published protocol (Ardisson-Araujo *et al.*, 2014b). For endonuclease restriction analyses, OBs were dissolved in alkaline solution and used to extract DNA (O'Reilly *et al.*, 1992). Both quantity and quality of the purified DNA were determined by electrophoresis on a 0.8% agarose gel (data not shown). The viral DNA (1–2 µg) was individually cleaved with the restriction enzymes *ApaI, BamHI, PstI, XbaI, XhoI, BglII, NsiI, and ClaI (Promega)* according to manufacturer's instructions. The DNA fragments were resolved by 0.8% agarose gel electrophoresis (Sambrook & Russel, 2001), visualized, and photographed in AlphaImager® Mini (Alpha Innotech).

5.3. Genome sequencing, assembly, and annotation

PeluSNPV genomic DNA was sequenced with a 454 Genome Sequencer (GS) FLX™ Standard (Roche) at the 'Centro de Genômica de Alto Desempenho do Distrito Federal' (Center of High-Performance Genomic, Brasilia, Brazil). The genome was assembled *de novo* using Geneious 7.0 (Kearse *et al.*, 2012) and confirmed using restriction enzyme digestion profile. One homologous region with low coverage was amplified (PeluOrf-7 F GGG TCA TAC ATC GTA TCA CCA AGC G and Pelu-p74 R CAT CTT ATC GGT TGG CGT ACG TGA C), cloned into pCRII (Invitrogen), and sequenced by Sanger (GENEWIZ®, Inc., USA). The open reading frames (ORFs) that started with a methionine codon (ATG) and encoded polypeptides of at least 50 amino acids were identified with Genious 7.0 and annotated using BLASTP (Altschul *et al.*, 1997).

5.4. Phylogenetic analyses

For *Baculoviridae* phylogenetic analysis, a MAFFT alignment (Katoh *et al.*, 2002) was carried out with concatenated amino acid sequences of 37 baculoviral core genes from 73 baculovirus genomes publicly available (Table S2). A maximum likelihood tree was inferred using a MAFFT alignment, the Fast-tree method (Stamatakis *et al.*, 2008)and a Shimodaira-Hasegawa-like test(Anisimova *et al.*, 2011). Horizontal gene transfer (HGT) events were investigated using the same method described above. MAFFT alignments (available upon request) of 36 sequences (for the *cypo016*-like genes) and 88 sequences (for *dut*genes) of homologs were used with the multiple sequence alignment package T-Coffee(Notredame *et al.*, 2000). Both the tree for *cypo016*-like and *dut* gene

were transformed to a cladogram using FigTree v1.4.0 in order to archive clarity. All the alignments are available upon request.

5.5. Viruses and insect cell line

Spodoptera frugiperda (fall armyworm) (Sf9) cells (Alami *et al.*, 2003) were maintained at 27 ºC in TC-100 medium (Invitrogen), supplemented with 10% fetal bovine serum (FBS, Invitrogen), penicillin G (60 µg/ml), streptomycin sulfate (200 µg/ml), and amphotericin B (0.5 µg/ml). Recombinant AcMNPV-C6 were propagated in insect cell cultures and their titers determined by end-point dilution (O'Reilly *et al.*, 1992).

5.6. Gene amplification, shuttle vectors, and recombinant AcMNPV virus construction

Gene from PeluSNPV (*pelu112*) and ErelGV (*erel005*) were separately amplified using two set of primers (Pelu F - ACA ACAGAG CTC ATG AAG ACC TAC ATT TGT GGT AC and Pelu R - AAT AGC GGC CGC TTA AAA AGT AGA TCC GAA TC, Erel F - ACA ACAGAG CTC ATG AAG ACC TAC ATT TGC GGT ACG and Erel R - AAA CGC GG CCG CTT AAG AAG TAG ACC CGA ACC) in two reactions which contained 100 ng of the DNA-template (PeluSNPV or ErelGV genomes), 300 µM of dNTP mix (Fermentas, Pittsburgh, PA, USA), 0.4 µM of each set of primer pairs, 1 U of VENT Polymerase (New England Biolabs, Ipswich, MA, USA), and 1x of the supplied reaction buffer. The reactions were subjected to the following program: 95 ºC/2 min, 35 cycles of 95 ºC/ 30s, 55 ºC/30 s and 68 ºC/1 min with a final extension of 5 min at 68

ºC. The amplified fragments were digested with *Sac*I/*Not*I (New England Biolabs, Ipswich, MA, USA) and cloned into pFB-PG-H-ha-pA shuttle plasmid (a modified pFB-PG containing a SV40-polyA signal and the *Drosophila melanogaster* hsp70 promoter to drive the heterolog gene expression with a for-fusion-ha-tag before the restriction sites) (Ardisson-Araujo *et al.*, 2015a) and confirmed by restriction digestion and sequencing (GENEWIZ®, Inc., USA). The modified plasmids containing the heterologous genes were transformed into DH10-Bac cells (Invitrogen, Carlsbad, CA, USA) by heat shock (Sambrook & Russel, 2001). Recombinant bacmids were selected and confirmed by PCR following the manufacturer's instructions (Bac-to-Bac®, Baculovirus expression systems, Invitrogen, Carlsbad, CA, USA). One µg of each recombinant bacmid was transfected into Sf9 cells (10^6) using Lipofectin (Wu & Passarelli, 2010). The supernatant of seven day post-transfection cells containing the recombinant viruses were collected, amplified in Sf9 cells, and titered as previously described (O'Reilly *et al.*, 1992).

Sf9 cells (1×10^6) were seeded on coverslips in 35-mm-diameter culture dishes and infected at MOI of 10 with recombinant viruses. At 24 h p.i., the supernatant was removed and the cells were washed twice with PBS, pH 6.2, and fixed in 2.5% formaldehyde in PBS for 10 min at room temperature (RT). The fixed cells were washed three times in PBS for 5 min, followed by permeabilization in 0.1% NP-40 (Sigma) in PBS for 10 min at RT. Cells were washed three times in PBS for 5 min per wash before incubation with blocking solution (5% BSA, 0.3% Triton-100 in PBS) for 1 h at RT, followed by incubation with anti-HA (1:500) in PBS with 1% BSA, 0.3% Triton X-100 overnight at 4 C in a humid chamber.Cells were washed three times in blocking solution for 5 min each, followed by 1 h incubation with Alexa Fluor 594-

conjugated goat anti-rabbit antibody (1:1,000) in the dark at RT. Cells were washed three times for 5 min each in PBS, followed by incubation with DAPI (Invitrogen) solution according to the manufacture instructions in PBS for 15 min at RT. The cells were subsequently washed three times for 10 min each in PBS. Coverslips were mounted on a glass slide with Fluoromount-G (SouthernBiotech) and stored at 4 °C in the dark until examined with a Carl Zeiss LSM 5 Pascal Laser Scanning Confocal Microscope.

5.7. Virus growth curves and polyhedra production

For viral growth curve analyses, three independent Sf9 cell dishes (0.5 x 10^6 cells/35-mm-diameter dish) were infected (MOI of 0.01 $TCID_{50}$/cell) for 1 h and then washed twice with TC-100 medium and replenished with 2 ml of fresh TC-100 medium supplemented with 10% FBS. The supernatants of the infected cells were collected at various time points to determine titers by 50% tissue culture infective dose (TCID50) endpoint dilution assays (O'Reilly *et al.*, 1992) on Sf9 cells. For polyhedra production, three independent infections were separately performed in Sf9 cells at 80% confluency in cell culture flasks (75 cm^2) at MOI of 5 $TCID_{50}$/cell. Cell monolayers were incubated for 1 h with the virus inocula, washed twice with TC-100 medium, and replenished with 12 ml fresh TC-100 medium supplemented with 10% FBS. The cells and polyhedra released were collected at 120 h p.i. and purified according to O'Reilly et al., (O'Reilly *et al.*, 1992). The purified OBs were diluted in the same volume, homogenized by vortexing overnight at 200 rpm, and counted using a hemocytometer.

5.8. Immunoblotting

Protein samples were mixed with equal volumes of 2x protein loading buffer (0.25 M Tris-Cl, pH 6.8, 4% SDS, 20% glycerol, 10% 2-mercaptoethanol, and 0.02% bromophenol blue) and incubated at 100°C for 5 min. Samples were resolved by 15 or 12% SDS-PAGE, transferred onto polyvinylidenefluoride (PVDF) membrane (Millipore), and probed with (i) mouse monoclonal anti-hemagglutinin (anti-HA) antibody (Covance), (ii) mouse monoclonal anti-GP64 antibody (eBioscience), or (iii) mouse polyclonal anti-IE-1 antibody; this probing was followed by incubation with horseradish peroxidase-conjugated secondary antibodies (Sigma). Blots were developed using the SuperSignal West Pico chemiluminescent substrate (Pierce) and exposed to X-ray films.

5.9. Quantitative real-time PCR (Q-PCR)

To detect viral DNA replication in virus-infected cells, Q-PCR was performed as previously described. Sf9 cells (1.0×10^6 cells/35-mm-diameter dish) were infected in triplicate at MOI of 5 $TCID_{50}$/cell, and cells were collected at different time points. Total DNA was prepared with the Wizard genomic DNA purification kit (Promega) according to the protocol of the manufacturer. Purified DNA was quantified by optical density measurement. Q-PCR was performed with 10 ng DNA and Absolute Q-PCR SYBR green fluorescein mix (Thermo Scientific) according to the protocol of the manufacturer by using the same primers to amplify a 100-bp region of the AcMNPV *gp41* gene as described previously (Vanarsdall *et al.*, 2005). Standard DNA samples were used from purified AcMNPV BV DNA and serially diluted to 100, 10, 1, 0.1,

0.01, and 0.001 ng. Genomic equivalents of DNA samples were determined by extrapolation from standard curves. A melting-curve analysis of each amplified sample was carried out to check the specificity of each reaction. The results were analyzed using GraphPad Prism version 5.01 (GraphPad Software, Inc.).

5.10. Homology modeling

The templates for three dimensional (3D) structure prediction of Pelu112 protein were searched in expasy SWISS-MODEL server (Biasini *et al.*, 2014) using the amino acid (aa) sequence as the reference. The Suitable templates were aligned with Pelu112 protein using T-Coffee server(Notredame *et al.*, 2000) and the resulting alignments were manually improved using BioEdit (Hall, 1999). Aligned sequences were used with MODELLERv9.10 (Sali & Blundell, 1993) to develop high quality 3D models. The highest quality models were selected and the accuracy of these predicted models was further analyzed through MolProbity (Chen *et al.*, 2010). The validation of all these models was done by checking the psi/phi ratio of Ramachandran plot obtained from MolProbity analysis. Yasara (Krieger *et al.*, 2009) was also applied for final models to check for energy minimization criteria. Ramachandran outlier residues were fixed with COOT (Emsley *et al.*, 2010) and energy minimization. The models were visualized usingThe PyMOL molecular graphics systemversion 1.0 (DeLano Scientific, San Carlos, CA).

6. References

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

Alami, M., Vacher, H., Bosmans, F., Devaux, C., Rosso, J. P., Bougis, P. E., Tytgat, J., Darbon, H. & Martin-Eauclaire, M. F. (2003). Characterization of Amm VIII from Androctonus mauretanicus mauretanicus: a new scorpion toxin that discriminates between neuronal and skeletal sodium channels. *Biochem J* **375**, 551-560.

Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25**, 3389-3402.

Anisimova, M., Gil, M., Dufayard, J. F., Dessimoz, C. & Gascuel, O. (2011). Survey of branch support methods demonstrates accuracy, power, and robustness of fast likelihood-based approximation schemes. *Syst Biol* **60**, 685-699.

Ardisson-Araujo, D. M., de Melo, F. L., Andrade, M. D., Sihler, W., Bao, S. N., Ribeiro, B. M. & de Souza, M. L. (2014a). Genome sequence of Erinnyis ello granulovirus (ErelGV), a natural cassava hornworm pesticide and the first sequenced sphingid-infecting betabaculovirus. *BMC Genomics* **15**, 856.

Ardisson-Araujo, D. M., Melo, F. L., de Souza Andrade, M., Brancalhao, R. M., Bao, S. N. & Ribeiro, B. M. (2014b). Complete genome sequence of the first non-Asian isolate of Bombyx mori nucleopolyhedrovirus. *Virus Genes*.

Ardisson-Araujo, D. M., Rohrmann, G. F., Ribeiro, B. M. & Clem, R. J. (2015a). Functional characterization of hesp018, a baculovirus-encoded serpin gene. *J Gen Virol* **96**, 1150-1160.

Ardisson-Araujo, D. M., Sosa-Gómez, D. R., Melo, F. L., Báo, S. N. & Ribeiro, B. M. (2015b). CHARACTERIZATION OF HELICOVERPA ZEA SINGLE NUCLEOPOLYHEDROVIRUS ISOLATED IN BRAZIL DURING THE FIRST OLD WORLD BOLLWORM (NOCTUIDAE: HELICOVERPA ARMIGERA) NATIONWIDE OUTBREAK. *Virus Reviews and Research* **20**, 4.

Becker, Y. (2000). Evolution of viruses by acquisition of cellular RNA or DNA nucleotide sequences and genes: an introduction. *Virus Genes* **21**, 7-12.

Biasini, M., Bienert, S., Waterhouse, A., Arnold, K., Studer, G., Schmidt, T., Kiefer, F., Cassarino, T. G., Bertoni, M., Bordoli, L. & Schwede, T. (2014). SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Res* **42**, W252-258.

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

Caradonna, S. J. & Cheng, Y. C. (1981). Induction of uracil-DNA glycosylase and dUTP nucleotidohydrolase activity in herpes simplex virus-infected human cells. *J Biol Chem* **256**, 9834-9837.

Chen, R., Wang, H. & Mansky, L. M. (2002). Roles of uracil-DNA glycosylase and dUTPase in virus replication. *J Gen Virol* **83**, 2339-2345.

Chen, V. B., Arendall, W. B., 3rd, Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S. & Richardson, D. C. (2010). MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr* **66**, 12-21.

Craveiro, S. R., Inglis, P. W., Togawa, R. C., Grynberg, P., Melo, F. L., Ribeiro, Z. M., Ribeiro, B. M., Bao, S. N. & Castro, M. E. (2015). The genome sequence of Pseudoplusia includens single nucleopolyhedrovirus and an analysis of p26 gene evolution in the baculoviruses. *BMC Genomics* **16**, 127.

Dube, D. K., Kunkel, T. A., Seal, G. & Loeb, L. A. (1979). Distinctive properties of mammalian DNA polymerases. *Biochim Biophys Acta* **561**, 369-382.

Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. (2010). Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* **66**, 486-501.

Escasa, S. R., Lauzon, H. A., Mathur, A. C., Krell, P. J. & Arif, B. M. (2006). Sequence analysis of the Choristoneura occidentalis granulovirus genome. *J Gen Virol* **87**, 1917-1933.

Ferrelli, M. L., Salvador, R., Biedma, M. E., Berretta, M. F., Haase, S., Sciocco-Cap, A., Ghiringhelli, P. D. & Romanowski, V. (2012). Genome of Epinotia aporema granulovirus

(EpapGV), a polyorganotropic fast killing betabaculovirus with a novel thymidylate kinase gene. *BMC Genomics* **13**, 14.

Goldberg, A. V., Romanowski, V., Federici, B. A. & Sciocco de Cap, A. (2002). Effects of the Epap granulovirus on its host, Epinotia aporema (Lepidoptera: Tortricidae). *J Invertebr Pathol* **80**, 148-159.

Hall, T. (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp Ser* **41**, 4.

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

Hinds, M. G., Norton, R. S., Vaux, D. L. & Day, C. L. (1999). Solution structure of a baculoviral inhibitor of apoptosis (IAP) repeat. *Nat Struct Biol* **6**, 648-651.

Huang, M. & Elledge, S. J. (1997). Identification of RNR4, encoding a second essential small subunit of ribonucleotide reductase in Saccharomyces cerevisiae. *Mol Cell Biol* **17**, 6105-6113.

Hughes, S. J., Johnston, L. H., de Carlos, A. & Smith, G. L. (1991). Vaccinia virus encodes an active thymidylate kinase that complements a cdc8 mutant of Saccharomyces cerevisiae. *J Biol Chem* **266**, 20103-20109.

Jehle, J. A., Blissard, G. W., Bonning, B. C., Cory, J. S., Herniou, E. A., Rohrmann, G. F., Theilmann, D. A., Thiem, S. M. & Vlak, J. M. (2006). On the classification and nomenclature of baculoviruses: a proposal for revision. *Arch Virol* **151**, 1257-1266.

Kato, A., Arii, J., Koyanagi, Y. & Kawaguchi, Y. (2015). Phosphorylation of Herpes Simplex Virus 1 dUTPase Regulates Viral Virulence and Genome Integrity by Compensating for Low Cellular dUTPase Activity in the Central Nervous System. *J Virol* **89**, 241-248.

Katoh, K., Misawa, K., Kuma, K. & Miyata, T. (2002). MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res* **30**, 3059-3066.

Katsuma, S., Kawaoka, S., Mita, K. & Shimada, T. (2008). Genome-wide survey for baculoviral host homologs using the Bombyx genome sequence. *Insect Biochem Mol Biol* **38**, 1080-1086.

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Meintjes, P. & Drummond, A. (2012). Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28**, 1647-1649.

Krieger, E., Joo, K., Lee, J., Raman, S., Thompson, J., Tyka, M., Baker, D. & Karplus, K. (2009). Improving physical realism, stereochemistry, and side-chain accuracy in homology modeling: Four approaches that performed well in CASP8. *Proteins* **77 Suppl 9**, 114-122.

Lacey, L. A., Hoffmann, D. F. & Federici, B. A. (2011). Histopathology and effect on development of the PhopGV on larvae of the potato tubermoth, Phthorimaea operculella (Lepidoptera: Gelechiidae). *J Invertebr Pathol* **108**, 52-55.

Ladner, R. D. (2001). The role of dUTPase and uracil-DNA repair in cancer chemotherapy. *Curr Protein Pept Sci* **2**, 361-370.

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

Lerner, D. L., Wagaman, P. C., Phillips, T. R., Prospero-Garcia, O., Henriksen, S. J., Fox, H. S., Bloom, F. E. & Elder, J. H. (1995). Increased mutation frequency of feline immunodeficiency virus lacking functional deoxyuridine-triphosphatase. *Proc Natl Acad Sci U S A* **92**, 7480-7484.

Lin, T., Chang, R., Wei, C. & Lang, G. (2012). Characterization of the deoxyuridine triphosphatase gene of Ophiusa disjungens nucleopolyhedrovirus. *Acta Virol* **56**, 241-246.

Lynn, D. E. & Hink, W. F. (1978). Cell cycle analysis and synchronization of the TN-368 insect cell line. *In Vitro* **14**, 236-238.

Miele, S. A., Garavaglia, M. J., Belaich, M. N. & Ghiringhelli, P. D. (2011). Baculovirus: molecular insights on their diversity and conservation. *Int J Evol Biol* **2011**, 379424.

Monier, A., Claverie, J. M. & Ogata, H. (2007). Horizontal gene transfer and nucleotide compositional anomaly in large DNA viruses. *BMC Genomics* **8**, 456.

Mustafi, D., Bekesi, A., Vertessy, B. G. & Makinen, M. W. (2003). Catalytic and structural role of the metal ion in dUTP pyrophosphatase. *Proc Natl Acad Sci U S A* **100**, 5670-5675.

Nardozzi, J. D., Lott, K. & Cingolani, G. (2010). Phosphorylation meets nuclear import: a review. *Cell Commun Signal* **8**, 32.

Notredame, C., Higgins, D. G. & Heringa, J. (2000). T-Coffee: A novel method for fast and accurate multiple sequence alignment. *J Mol Biol* **302**, 205-217.

O'Reilly, D., Miller, L. K. & Luckow, V. A. (1992). *Baculovirus Expression Vectors: a laboratory manual*. New York: Freeman and Company.

Oliveira, J. V., Wolff, J. L., Garcia-Maruniak, A., Ribeiro, B. M., de Castro, M. E., de Souza, M. L., Moscardi, F., Maruniak, J. E. & Zanotto, P. M. (2006). Genome of the most widely used viral biopesticide: Anticarsia gemmatalis multiple nucleopolyhedrovirus. *J Gen Virol* **87**, 3233-3250.

Oliveros, M., Garcia-Escudero, R., Alejo, A., Vinuela, E., Salas, M. L. & Salas, J. (1999). African swine fever virus dUTPase is a highly specific enzyme required for efficient replication in swine macrophages. *J Virol* **73**, 8934-8943.

Pardo, E. G. & Gutierrez, C. (1990). Cell cycle- and differentiation stage-dependent variation of dUTPase activity in higher plant cells. *Exp Cell Res* **186**, 90-98.

Penades, J. R., Donderis, J., Garcia-Caballer, M., Tormo-Mas, M. A. & Marina, A. (2013). dUTPases, the unexplored family of signalling molecules. *Curr Opin Microbiol* **16**, 163-170.

Priet, S., Sire, J. & Querat, G. (2006). Uracils as a cellular weapon against viruses and mechanisms of viral escape. *Curr HIV Res* **4**, 31-42.

Primo, L. M., Duarte, J. A. & Machado, I. C. (2013). Hawkmoth fauna (Sphingidae, Lepidoptera) in a semi-deciduous rainforest remnant: composition, temporal fluctuations, and new records for Northeastern Brazil. *An Acad Bras Cienc* **85**, 1177-1188.

Pyles, R. B., Sawtell, N. M. & Thompson, R. L. (1992). Herpes simplex virus type 1 dUTPase mutants are attenuated for neurovirulence, neuroinvasiveness, and reactivation from latency. *J Virol* **66**, 6706-6713.

Reichard, P. (1988). Interactions between deoxyribonucleotide and DNA synthesis. *Annu Rev Biochem* **57**, 349-374.

Rohrmann, G. F. (2013). *Baculovirus Molecular Biology*. Bethesda (MD): National Center for Biotechnology Information (US). Available from: http://www.ncbi.nlm.nih.gov/books/NBK49500/.

Ross, J., Williams, M. & Cohen, J. I. (1997). Disruption of the varicella-zoster virus dUTPase and the adjacent ORF9A gene results in impaired growth and reduced syncytia formation in vitro. *Virology* **234**, 186-195.

Sali, A. & Blundell, T. L. (1993). Comparative protein modelling by satisfaction of spatial restraints. *J Mol Biol* **234**, 779-815.

Sambrook, J. & Russel, D. W. (2001). *Molecular Cloning : a laboratory manual*. New York: Cold Spring Harbor.

Sosa-Gómez, D. R., Kitajima, E. W. & Rolon, M. E. (1994). First records of entomopathogenic diseases in the paraguay tea agroecosystem in Argentina. *Florida Entomologist* **77**, 5.

Stamatakis, A., Hoover, P. & Rougemont, J. (2008). A rapid bootstrap algorithm for the RAxML Web servers. *Syst Biol* **57**, 758-771.

Steagall, W. K., Robek, M. D., Perry, S. T., Fuller, F. J. & Payne, S. L. (1995). Incorporation of uracil into viral DNA correlates with reduced replication of EIAV in macrophages. *Virology* **210**, 302-313.

Strahler, J. R., Zhu, X. X., Hora, N., Wang, Y. K., Andrews, P. C., Roseman, N. A., Neel, J. V., Turka, L. & Hanash, S. M. (1993). Maturation stage and proliferation-dependent expression of dUTPase in human T cells. *Proc Natl Acad Sci U S A* **90**, 4991-4995.

Threadgill, D. S., Steagall, W. K., Flaherty, M. T., Fuller, F. J., Perry, S. T., Rushlow, K. E., Le Grice, S. F. & Payne, S. L. (1993). Characterization of equine infectious anemia virus dUTPase: growth properties of a dUTPase-deficient mutant. *J Virol* **67**, 2592-2600.

Tsai, M. F., Yu, H. T., Tzeng, H. F., Leu, J. H., Chou, C. M., Huang, C. J., Wang, C. H., Lin, J. Y., Kou, G. H. & Lo, C. F. (2000). Identification and characterization of a shrimp white spot syndrome virus (WSSV) gene that encodes a novel chimeric polypeptide of cellular-type thymidine kinase and thymidylate kinase. *Virology* **277**, 100-110.

Turelli, P., Guiguen, F., Mornex, J. F., Vigne, R. & Querat, G. (1997). dUTPase-minus caprine arthritis-encephalitis virus is attenuated for pathogenesis and accumulates G-to-A substitutions. *J Virol* **71**, 4522-4530.

Turelli, P., Petursson, G., Guiguen, F., Mornex, J. F., Vigne, R. & Querat, G. (1996). Replication properties of dUTPase-deficient mutants of caprine and ovine lentiviruses. *J Virol* **70**, 1213-1217.

Tzeng, H. F., Chang, Z. F., Peng, S. E., Wang, C. H., Lin, J. Y., Kou, G. H. & Lo, C. F. (2002). Chimeric polypeptide of thymidine kinase and thymidylate kinase of shrimp white spot syndrome virus: thymidine kinase activity of the recombinant protein expressed in a baculovirus/insect cell system. *Virology* **299**, 248-255.

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

Vanarsdall, A. L., Okano, K. & Rohrmann, G. F. (2005). Characterization of the replication of a baculovirus mutant lacking the DNA polymerase gene. *Virology* **331**, 175-180.

Warner, H. R., Thompson, R. B., Mozer, T. J. & Duncan, B. K. (1979). The properties of a bacteriophage T5 mutant unable to induce deoxyuridine 5'-triphosphate nucleotidohydrolase. Synthesis of uracil-containing T5 deoxyribonucleic acid. *J Biol Chem* **254**, 7534-7539.

Williams, M. V. (1988). Herpes simplex virus-induced dUTPase: target site for antiviral chemotherapy. *Virology* **166**, 262-264.

Wu, W. & Passarelli, A. L. (2010). Autographa californica multiple nucleopolyhedrovirus Ac92 (ORF92, P33) is required for budded virus production and multiply enveloped occlusion-derived virus formation. *J Virol* **84**, 12351-12361.

Xu, P., Liu, Y., Graham, R. I., Wilson, K. & Wu, K. (2014). Densovirus Is a Mutualistic Symbiont of a Global Crop Pest (Helicoverpa armigera) and Protects against a Baculovirus and Bt Biopesticide. *PLoS Pathog* **10**, e1004490.

Yan, H. & Tsai, M. D. (1999). Nucleoside monophosphate kinases: structure, mechanism, and substrate specificity. *Adv Enzymol Relat Areas Mol Biol* **73**, 103-134, x.

7. SUPPLEMENTARY MATERIAL

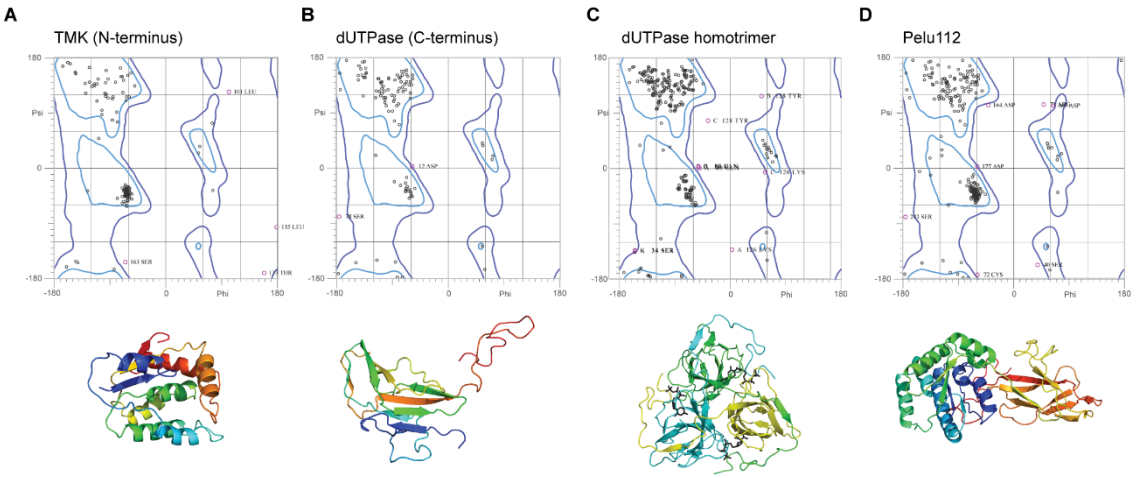

Figure S1. Ramachandran plot for each protein model proposed.(A) N-terminal region of Pelu112, the TMK-like enzyme. (B) C-terminal region of Pelu112, the dUTPase-like enzyme. (C) dUTPase homotrimer. (C) Fused model. The individual, assembled, or fused structures are shown below each respective plot.

Table S1. Characteristics of the Perigonia lusca single nucleopolyhedrovirus (PeluSNPV) genome: analysis and homology search. Predicted ORFs are compared with homolog genes in two related genomes.

Orf	Name	Position	Size (nt)	Size (aa)	ClbiNPV		AcMNPV		Best hit
					Orf	Max Id (%)[+]	Orf	Max Id (%)[+]	
1	*polh*	1 > 741	741	246	1	91	8	89	OrleNPV
2	*orf1629*	902 < 2,566	1,665	554	2	27	9	29	ClbiNPV
3	*pk-1*	2,559 > 3,356	798	265	3	55	10	44	AgseNPV-B
4[a]		3,629 < 4,042	414	137	-	-	-	-	*Ceriporiopsis subvermispora*
5	*hoar*	4,128 < 6,239	2,112	703	4	30			EcobNPV
6[a]		6,469 > 6,657	189	62	-	-	-	-	*Daphnia pulex*
7[a]		6,804 > 8,276	1,473	490	-	-	-	-	*Megasphaera sp.*
	hr1	8,291 - 9,364	1,074	-	-	-	-	-	-
8	*p74*	9,382 < 11,358	1,977	658	14	62	138	59	OrleNPV
9	*me53*	11,452 < 12,522	1,071	356	12	49	139	23	ClbiNPV
10[a]		12,561 > 12,713	153	50	-	-	-	-	*Beta vulgaris*
11	*ie-0*	12,860 > 13,693	834	277	11	40	141	28	ChchNPV
12	*p49*	13,742 > 15,265	1,524	507	10	71	142	51	ClbiNPV
13	*odv-e18*	15,210 > 15,470	261	86	10b*	71[§]	143	83	LyxyMNPV
14	*odv-e27*	15,508 > 16,371	864	287	9	67	144	49	OrleNPV

15	*chtb-1*	16,388	>	16,669	282	93	9b*	69[§]	145	49	AdhoNPV
16	*ep23*	16,680	<	17,300	621	206	8	34	146	33	ApciNPV
17 [a]		17,380	>	17,838	459	152	-	-	-	-	*no hit*
18 [a]		17,901	>	18,380	480	159	-	-	-	-	no hit
19	*ie-1*	18,245	>	19,615	1,371	456	7	41	147	31	EcobNPV
20	*odv-e56 (pif-5)*	19,769	<	20,836	1,068	355	6	61	148	54	OrleNPV
	hr2	20,849	-	21,997	1,149	-	-	-	-	-	-
21	*p47*	22,004	>	23,218	1,215	404	36	65	40	54	HespNPV
22	*dbp-1*	23,378	<	24,301	924	307	27	46	25	29	ClbiNPV
23	*nudix; bv-e31*	24,479	>	25,354	876	291	25	57	38	52	AgipMNPV
24	*lef-11*	25,198	>	25,800	603	200	-	-	37	34	AgseNPV-B
25	*39k*	25,736	>	26,689	954	317	24	43	36	37	ClbiNPV
26 [a]		26,833	<	27,012	180	59	-	-	-	-	no hit
27	*v-ubq*	26,969	<	27,247	279	92	22	80	35	76	HespNPV
28	*lef-7*	27,414	>	28,388	975	324	37	39	125	31	MaviNPV
29		28,404	>	29,021	618	205	21	54	34	33	HespNPV
30	*p10*	29,152	<	29,415	264	87	20	63	137	29	ChchNPV
31	*p26-1*	29,517	<	30,377	861	286	19	42	136	32	ApciNPV
32		30,583	>	30,834	252	83	18	48	29	29	MacoNPV-A
33	*lef-6*	30,963	<	31,733	771	256	17	55	28	47	AgseNPV

34	*dbp-2*	31,781	<	32,560	780	259	16	36	25	30	AgseNPV
35 [a]		32,780	>	32,938	159	52	-	-	-	-	*Saccharomonospora viridis*
36	*lef-12*	33,316	>	34,035	720	239	34	28	41	37	AdorNPV
37		34,025	>	34,276	252	83	33	38	43	31	BomaNPV
38		34,295	<	34,843	549	182	-	-	-	-	MacoNPV-A
39	*ctl-1*	34,946	>	35,131	186	61	53	60	3	40	ChmuNPV
40	*lef-9*	35,207	<	36,703	1,497	498	47	76	62	68	ChchNPV
41	*fp-25k*	36,892	>	37,536	645	214	46	67	61	60	OrleNPV
42	*bro-a*	37,725	>	38,108	384	127	105	52	-	-	HespNPV
43	*chab-a*	38,202	>	38,516	315	104	45	56	60	56	LdMNPV
44	*chab-b*	38,568	>	39,068	501	166	44	72	58/59	47	MacoNPV-A
45		39,121	<	39,693	573	190	43	42	57	43	BusuNPV
46		40,084	<	40,338	255	84	-	-	-	-	AdorNPV
47		40,277	<	40,564	288	95	41	59	55	40	SujuNPV
48 [a]		40,545	>	40,751	207	68	-	-	-	-	*no hit*
49	*vp1054*	40,685	<	41,749	1,065	354	39	48	54	39	AgipMNPV
50	*lef-10*	41,604	<	41,837	234	77	-	-	53a	33	TnSNPV
51		41,800	>	42,030	231	76	-	-	-	-	TnSNPV
52		42,047	>	43,141	1,095	364	38	31	-	-	HespNPV
53		43,130	<	43,555	426	141	28	63	53	46	OrleNPV

54 [a]		43,613	>	43,858	246	81	-	-	-	-	*no hit*
55 [a]		43,708	<	43,965	258	85	-	-	-	-	*Flavobacterium soli*
56	*dnaj*	43,990	<	44,886	897	298	31	34	-	-	ClbiNPV
57	*lef-8*	44,907	>	47,582	2,676	891	32	69	50	62	ApciNPV
58	*gp37*	47,791	<	48,741	951	316	56	57	64	47	ClbiNPV
59		48,918	<	49,124	207	68	58	46	111	52	BmNPV
60	*chitinase*	49,277	<	50,989	1,713	570	59	72	126	71	ClbiNPV
61	*v-cath*	51,109	>	52,122	1,014	337	60	69	127	69	SujuNPV
62	*p26-2*	52,171	<	52,899	729	242	61	44	136	28	ClbiNPV
63	*chtB-2*	52,992	<	53,330	339	112	62	45	150	31	HaNPV
64	*iap-2*	53,334	<	54,077	744	247	63	34	71	31	AgseNPV
65	*mtase-1*	54,074	<	54,886	813	270	64	52	69	49	SpliNPV-II
66		54,858	<	55,232	375	124	-	-	68	41	AgseNPV
67	*lef-3*	55,394	>	56,470	1,077	358	65	43	67	28	ClbiNPV
68	*desmoplakin*	56,641	<	58,983	2,343	780	66	32	66	31	TnSNPV
69	*dna-pol*	59,021	>	62,206	3,186	1061	67	65	65	48	ClbiNPV
70		62,300	<	62,689	390	129	68	60	75	31	ClbiNPV
71		62,697	<	62,954	258	85	69	72	76	59	OrleNPV
72	*vlf-1*	63,029	<	64,207	1,179	392	71	81	77	73	ClbiNPV
73		64,219	<	64,569	351	116	72	72	78	59	BusuNPV

74	*gp41*	64,640	<	65,602	963	320	73	79	80	60	ClbiNPV
75		65,729	<	66,247	519	172	74	57	81	60	TnSNPV
76	*tlp20*	66,177	<	66,938	762	253	75	48	82	35	EupsNPV
77	*p95 (vp91)*	66,808	>	69,255	2,448	815	76	39	83	35	ApciNPV
78	*cg30*	69,491	<	70,228	738	245	77	32	88	32	OrleNPV
79	*vp39*	70,330	<	71,337	1,008	335	78	58	89	40	ClbiNPV
80	*lef-4*	71,336	>	72,886	1,551	516	79	53	90	45	HespNPV
81	*p33 (sox)*	72,916	<	73,617	702	233	80	65	92	47	ClbiNPV
82	*p18*	73,696	>	74,196	501	166	81	66	93	48	PespNPV
83	*odv-e25*	74,193	>	74,888	696	231	82	72	94	42	ClbiNPV
84	*dna-helicase*	75,018	<	78,680	3,663	1220	83	58	95	42	OrleNPV
85	*odv-e28 (pif-4)*	78,649	>	79,173	525	174	84	61	96	50	OrleNPV
86	*38k*	79,214	<	80,254	1,041	346	85	59	98	49	ClbiNPV
87	*lef-5*	80,150	>	81,037	888	295	86	61	99	48	OrleNPV
88	*p6.9*	81,055	<	81,285	231	76	86b*	44[§]	100	42[§]	*no hit*
89[a]		81,240	>	81,410	171	56	-	-	-	-	*no hit*
90	*p40*	81,347	<	82,522	1,176	391	87	56	101	39	ClbiNPV
91	*p12*	82,541	<	82,912	372	123	88	59	102	36	ClbiNPV
92	*p48/p45*	82,905	<	84,092	1,188	395	89	69	103	39	ClbiNPV
93	*vp80*	84,121	>	86,733	2,613	870	90	28	104	24	ClbiNPV

94		86,755	>	86,922	168	55	91	59	110	35	EcobNPV
95	*odv-ec43*	86,929	>	88,008	1,080	359	92	72	109	43	ClbiNPV
96		88,077	>	88,367	291	96	-	-	-	-	SfMNPV
97	*p13*	88,397	<	89,218	822	273	94	62	-	-	SpliNPV-II
98		89,273	>	90,373	1,101	366	95	31	112/113	36	LyxyMNPV
99 [a]		90,556	>	90,906	351	116	-	-	-	-	*Arabidopsis thaliana*
100 [a]		90,810	<	91,187	378	125	-	-	-	-	no hit
101 [a]		91,056	>	91,829	774	257	-	-	-	-	*Halomonas sp.*
102	*iap-3*	91,830	>	92,432	603	200	-	-	-	-	LdMNPV
103		92,443	<	93,135	693	230	97	92	106	64	ClbiNPV
104		93,294	>	94,025	732	243	-	-	-	-	ErelGV
105	*pagr*	94,075	<	95,580	1,506	501	98	21	-	-	SujuNPV
106		95,671	<	96,060	390	129	99	39	-	-	ApciNPV
107	*pif-3*	96,071	<	96,697	627	208	100	44	115	44	SpliNPV-II
108	*sod*	96,777	>	97,268	492	163	102	76	31	73	ClbiNPV
109		97,317	<	98,336	1,020	339	-	-	11	47	AgMNPV
110		98,277	>	98,459	183	60	-	-	-	-	ChroNPV
111	*ctl-2*	98,483	>	98,644	162	53	-	-	3	74	AcMNPV
112	*dut-fused*	98,827	>	99,780	954	317	-	-	-	-	ErelGV
113		99,938	>	100,315	378	125	103	33	-	-	AgseNPV

114		100,312	>	100,590	279	92	104	40	117	40	ClbiNPV
115	*pif-2*	100,652	<	101,794	1,143	380	107	72	22	66	BusuNPV
116	*pkip*	101,837	<	102,415	579	192	108	33	-	-	HzSNPV
117	*lef-2*	102,469	<	103,098	630	209	109	57	6	42	ClbiNPV
118		103,070	<	103,438	369	122	110	43	-	-	AdorNPV
119 [a]		103,643	<	104,098	456	151	-	-	-	-	*Sulfolobus islandicus*
120 [a]		104,260	<	104,409	150	49	-	-	-	-	no hit
121	*p24*	104,462	>	105,241	780	259	111	54	129	40	HespNPV
122		105,242	<	105,712	471	156	112	30	-	-	HespNPV
123	*gp16*	105,810	>	106,106	297	98	113	49	130	37	TnSNPV
124	*he65*	106,230	>	107,057	828	275	-	-	105	35	AgseGV
125	*pep; pp34*	107,208	>	108,128	921	306	114	55	131	28	OrleNPV
126	*rr2a*	108,210	<	109,262	1,053	350	-	-	-	-	HespNPV
127		109,345	<	109,758	414	137	115	44	19	36	OrleNPV
128		109,769	>	110,989	1,221	406	116	32	18	27	AgseNPV
129	*alk-exo*	111,007	>	112,266	1,260	419	117	48	133	39	ApciNPV
130		112,341	<	113,060	720	239	-	-	-	-	AgseNPV-B
131	*fgf*	113,223	>	114,365	1,143	380	118	33	32	21	SujuNPV
132		114,379	<	114,615	237	78	-	-	-	-	AgseNPV-B
133	*pif-1*	114,618	<	116,240	1,623	540	120	48	119	51	ApciNPV

134	*odv-e66*	116,280	<	118,259	1,980	659	-	-	46	45	OrleNPV
135	*f protein*	118,381	<	120,507	2,127	708	129	67	23	23	ClbiNPV
136		120,663	>	123,596	2,934	977	128	42	-	-	ClbiNPV
137		123,633	<	124,490	858	285	127	33	17	36	HespNPV
138		124,598	<	125,272	675	224	126	48	-	-	ClbiNPV
139	*egt*	125,504	<	127,117	1,614	537	125	47	15	52	AcMNPV
140 [a]		127,192	<	127,401	210	69	-	-	-	-	no hit
141		127,362	<	127,700	339	112	124	56	-	-	OrleNPV
142	*lef-1*	127,719	>	128,411	693	230	123	48	14	42	ClbiNPV
143	*38.7k*	128,429	>	129,580	1,152	383	122	39	13	41	ClbiNPV
144 [a]		129,678	<	130,292	615	204	-	-	-	-	*Plasmodium vinckei petteri*
145	*rr1*	130,348	<	132,645	2,298	765	-	-	-	-	EupsNPV

+: identity obtained by BLASTX.
a: unique gene
*: not annotated in the Genbank database genome
§: acquired by manual alignment using the MAFFT method

Table S2. Species used in this paper for reconstruction of the baculovirus phylogeny in the FIG. 2. The species from the genera *Alphabaculovirus* from Group I (red) and Group II (black), *Betabaculovirus* (dark blue), *Gammabaculovirus* (orange), and *Deltabaculovirus* (light blue) are presented here together with abbreviation used in the main text, host family from where the virus was isolated, and the Genbank accession number.

Species	**Abbreviation**	**Host family**	**Accession**
Adoxophyes honmai nucleopolyhedrovirus	AdhoNPV	Tortricidae	AP006270
Adoxophyes orana nucleopolyhedrovirus	AdorNPV	Tortricidae	EU591746
Agrotis ipsilon multiple nucleopolyhedrovirus strain illinois	AgipMNPV	Noctuidae	EU839994
Agrotis segetum nucleopolyhedrovirus	AgseNPV	Noctuidae	DQ123841
Apocheima cinerarium nucleopolyhedrovirus	ApciNPV	Geometridae	FJ914221
Buzura suppressaria nucleopolyhedrovirus	BusuNPV	Geometridae	KF611977
Chrysodeixis chalcites nucleopolyhedrovirus	ChchNPV	Noctuidae	AY864330
Clanis bilineata nucleopolyhedrovirus	ClbiNPV	Sphingidae	DQ504428
Ectropis obliqua nucleopolyhedrovirus strain A1	EcobNPV-A1	Geometridae	DQ837165
Euproctis pseudoconspersa nucleopolyhedrovirus	EupsNPV	Lymantriidae	FJ227128
Helicoverpa armigera multiple nucleopolyhedrovirus	HaMNPV	Noctuidae	EU730893
Helicoverpa armigera nucleopolyhedrovirus C1	HaNPV-C1	Noctuidae	AF303045
Helicoverpa zea single nucleopolyhedrovirus USA	HzSNPV-USA	Noctuidae	AF334030
Hemileuca sp. nucleopolyhedrovirus	HespNPV	Saturniidae	KF158713
Lambdina fiscellaria nucleopolyhedrovirus	LafiNPV	Geometriidae	KP752043
Leucania separata nuclear polyhedrovirus strain AH1	LeseNPV	Noctuidae	AY394490
Lymantria díspar multiple nucleopolyhedrovirus	LdMNPV	Lymantriidae	AF081810
Lymantria xylina multiple nucleopolyhedrovirus	LyxyMNPV	Lymantriidae	GQ202541
Mamestra brassicae multiple nucleopolyhedrovirus strain Chb1	MbMNPV-CHb1	Noctuidae	JX138237

Mamestra configurata nucleopolyhedrovirus-A strain 90/2	MacoNPV-A 90/2	Noctuidae	U59461
Mamestra configurata nucleopolyhedrovirus B	MacoNPV-B	Noctuidae	AY126275
Orgyia leucostigma nucleopolyhedrovirus isolate CFS-77	OrleNPV	Lymantriidae	EU309041
Peridroma sp. nucleopolyhedrovirus	PespNPV	Noctuidae	KM009991
Perigonia lusca single nucleopolyhedrovirus	**PeluSNPV**	**Sphigidae**	**KM596836**
Pseudoplusia includens single nucleopolyhedrovirus IE	PsinSNPV	Noctuidae	KJ631622
Spodoptera exigua nucleopolyhedrovirus	SeMNPV	Noctuidae	AF169823
Spodoptera frugiperda multiple nucleopolyhedrovirus isolate 19	SfMNPV-19	Noctuidae	EU258200
Spodoptera litoralis nucleopolyhedrovirus isolate AN1956	SpliNPV-1956	Noctuidae	JX454574
Spodoptera litura nucleopolyhedrovirus G2	SpliNPV-G2	Noctuidae	AF325155
Spodoptera litura nucleopolyhedrovirus II	SpliNPV-II	Noctuidae	EU780426
Sucra jujuba nucleopolyhedrovirus	SujuNPV	Geometridae	KJ676450
Trichoplusia ni single nucleopolyhedrovirus	TnSNPV	Noctuidae	DQ017380
Autographa californica nucleopolyhedrovirus clone C6	AcMNPV-C6	Noctuidae	L22858
Anticarsia gemmatalis nucleopolyhedrovirus	AgMNPV	Noctuidae	DQ813662
Antheraea pernyi nucleopolyhedrovirus isolate L2	AnpeNPV-L2	Saturniidae	EF207986
Bombyx mori nucleopolyhedrovirus strain T3	BmNPV-T3	Bombycidae	L33180
Bombyx mandarina nucleopolyhedrovirus S2	BomaNPV-S2	Bombycidae	JQ071499
Choristoneura fumiferana defective multiple nucleopolyhedrovirus	CfDEFMNPV	Tortricidae	AY327402
Choristoneura fumiferana multiple nucleopolyhedrovirus	CfMNPV	Tortricidae	AF512031
Choristoneura murinana nucleopolyhedrovirus	ChmuNPV	Tortricidae	KF894742
Choristoneura occidentalis nucleopolyhedrovirus	ChocNPV	Tortricidae	KC961303
Choristoneura rosaceana nucleopolyhedrovirus	ChroNPV	Tortricidae	KC961304
Condylorrhiza vestigialis multiple nucleopolyhedrovirus	CoveMNPV	Crambidae	KJ631623

Dendrolimus kikuchii nucleopolyhedrovirus	DekiNPV	Lasiocampidae	JX193905
Epiphyas postvittana nucleopolyhedrovirus	EppoNPV	Tortricidae	AY043265
Hyphantria cunea nucleopolyhedrovirus	HycuNPV	Arctiidae	AP009046
Maruca vitrata multiple nucleopolyhedrovirus	MaviMNPV	Crambidae	EF125867
Orgyia pseudotsugata multiple nucleopolyhedrovirus	OpMNPV	Lymantriidae	U75930
Philosamia cynthia ricini nucleopolyhedrovirus	PhcyNPV	Saturniidae	JX404026
Plutella xylostella multiple nucleopolyhedrovirus isolate CL3	PlxyMNPV	Plutellidae	DQ457003
Rachiplusia ou multiple nucleopolyhedrovirus	RoMNPV	Noctuidae	AY145471
Thysanoplusia orichalcea nucleopolyhedrovirus	ThorNPV	Noctuidae	JX467702
Adoxophyes orana granulovirus	AdorGV	Tortricidae	AF547984
Agrotis segetum granulovirus-L1	AgseGV-L1	Noctuidae	KC994902
Choristoneura occidentalis granulovirus	ChocGV	Tortricidae	DQ333351
Clostera anastomosis granulovirus	CaLGV	Notodontidae	KC179784
Clostera anachoreta granulovirus	ClanGV	Notodontidae	HQ116624
Cryptophlebia leucotreta granulovirus isolate CV3	CrleGV	Tortricidae	AY229987
Cydia pomonella granulovirus	CpGV	Tortricidae	U53466
Epinotia aporema granulovirus	EpapGV	Tortricidae	JN408834
Erinnyis ello granulovirus	ErelGV	Sphingidae	KJ406702
Helicoverpa armigera granulovirus	HaGV	Noctuidae	EU255577
Phthorimaea operculella granulovirus	PhopGV	Gelechiidae	AF499596
Pieris rapae granulovirus E3	PiraGV-E3	Pieridae	GU111736
Plutella xylostella granulovirus	PlxyGV	Plutellidae	AF270937
Pseudaletia unipuncta granulovirus	PsunGV-Hawaiin	Noctuidae	EU678671
Spodoptera frugiperda granulovirus	SpfrGV	Noctuidae	KM371112

Spodoptera litura granulovirus isolate K1	SpliGV	Noctuidae	DQ288858
Xestia c-nigrum granulovirus	XcGV	Noctuidae	AF162221
Neodiprion sertifer nucleopolyhedrovirus	NeseNPV	Diprionidae	AY430810
Neodiprion lecontei nucleopolyhedrovirus	NeleNPV	Diprionidae	AY349019
Neodiprion abietis nucleopolyhedrovirus	NeabNPV	Diprionidae	DQ317692
Culex nigripalpus nucleopolyhedrovirus	CuniNPV	Culicidae	AF403738

Capítulo 9. Discussão Geral

O interesse pelo estudo de doenças associadas a insetos tem seu início num passado bastante remoto com a primeira descrição formal da ‘wilting disease’ (do inglês, doença do murchamento) acometendo larvas do bicho da seda (*Bombyx mori*) no século XVI (Herniou et al., 2003). Somente em meados do século XX foi observada uma partícula viral com a forma de bastão associada a tal doença, característico da família *Baculoviridae*. Contrapondo-se a insetos benéficos, muitas espécies são consideradas pragas no contexto de interação com humanos ao competirem por alimentos cultivados. Felizmente, tais populações são susceptíveis a infecções virais, o que impulsiona também o estudo de virologia de insetos como agentes para o controle biológico.

Várias famílias virais de insetos foram descritas e com o advento e diminuição de custos de sequenciamento em larga escala, o número de espécies cresce vigorosamente. De fato, para a maioria desses vírus, pouco se é sabido da evolução, de aspectos moleculares da infecção, da interação com o hospedeiro e do papel na dinâmica de população dos hospedeiros. Dessa forma, neste trabalho, vários genomas de baculovírus isolados no Brasil foram sequenciados e descritos: betabaculovírus das espécies *Erinnyis ello granulovirus* (Capítulo 3) e *Diatraea saccharalis granulovirus* (Capítulo 6) e alphabaculovirus das espécies *Bombyx mori nucleopolyhedrovirus* (Capítulo 2), *Helicoverpa zea single nucleopolyhedrovirus* (Capítulo 4) e *Perigonia lusca single nucleopolyhedrovirus* (Capítulo 8). Concomitante à descrição do genoma, caracterizamos estruturalmente algumas espécies, avaliamos a taxa de mortalidade em situações controladas de infecção, bem como caracterizamos alguns genes que permitiram um entendimento evolutivo mais amplo das espécies descritas e de sua

interação com o hospedeiro. Concernente ao estudo de baculovírus, o estudo da interação patógeno-hospedeiro e sua evolução pode ser estendida para organismos relacionados como outros vírus de DNA dupla-fita ou vírus oralmente infectivos associados a insetos, como arboviroses.

Aquisições por transferência horizontal, perdas e duplicações gênicas são as principais forças que dirigem a diversificação de baculovírus e refletem a natureza fluídica de seu genoma (Herniou et al., 2001). Contudo, tanto a troca de informações quanto sua fixação no genoma ocorre por mecanismos moleculares desconhecidos, apesar de recorrentes em regiões de alta repetição de bases (Capítulo 5) (Ardisson-Araujo et al., 2015). Organismos fontes de genes incluem não somente o inseto hospedeiro e outras espécies de baculovírus, como também bactérias, plantas e outras famílias virais (Ardisson-Araujo et al., 2015; de Castro Oliveira et al., 2008; Kamita et al., 2005; Theze et al., 2015). Essa troca pode estar relacionada ao fato de que vírus de insetos com diferentes origens filogenéticas exploram o mesmo nicho ecológico. Assim, pressão de seleção similar tende a forçar os organismos a evoluírem adaptações convergentes como semelhança de conteúdo genômico mediado por aquisição de genes e compartilhamento (Theze et al., 2015). Por exemplo, genes relacionados a metabolismo de nucleotídeo parecem ter sido adquiridos de forma independente pelo menos nove vezes em baculovírus.

Mecanismo de intercâmbio gênico em alguns contextos pode ser de certa forma, intuitivo como quando associado a diferentes espécies de baculovírus, uma vez que tanto a molécula alvo quanto a molécula doadora apresentam semelhante composição bioquímica. Por exemplo, aquisições independentes de um mesmo gene relacionado a

metabolismo de nucleotídeo foram observadas no genoma de espécies de baculovírus distantemente relacionadas (*dut-tmk* no Capítulo 8). Dessa forma, eventos de co-infecção de uma mesma célula hospedeira poderiam desenhar o cenário ideal de troca de informação por recombinação gênica clássica. Vários patógenos intracelulares estritos apresentam mecanismos que evitam co-infecções (Beperet et al., 2014). Entretanto, no caso de baculovírus, o conteúdo gênico aponta eventos de co-infecção como recorrentes. Alphabaculovírus e betabaculovírus podem explorar diferentes receptores para entrada na célula, uma vez que a proteína de envelope do vírus responsável pelo espalhamento da infecção varia (Westenberg et al., 2007). Isso poderia explicar o intenso fluxo gênico entre alphabaculovírus e betabaculovírus (Cuartas et al., 2015), uma vez que não haveria competição direta de receptores para entrada na célula e estabelecimento de um cenário de co-infecção.

O mecanismo de troca de informações entre baculovírus seria possivelmente estendido para outros vírus de inseto da classe I e II, isto é com genoma de dsDNA e ssDNA com intermediáio dsDNA, como é o caso de entomopoxvirus, iridovirus; vírus gigantes e densovírus (Capítulo 6). Densovírus são vírus de DNA fita-simples capazes de infectar diferentes ordens de inseto causando doença ou protegendo ao hospedeiro ao qual está associado (Xu et al., 2014). No genoma de ErelGV, PeluSNPV e DisaGV foram encontrados genes associados à proteína não-estrutural de densovírus (Capítulo 3, 6 e 8). Uma vez que intermediários de replicação de vírus de ssDNA apresentam dsDNA, é razoável pensar que este pode ter sido adquirido e fixado num evento de recombinação com o intermediário replicativo. Por outro lado, para classes virais distintas, os mecanismos de recombinação gênica se tornam obscuros. Por exemplo, aquisição de genes do hospedeiro (*e.g. serpin*, Capítulo 5) ou de vírus com genoma de RNA (*e.g.* o

doador da proteína de envelope *gp64*, Capítulos 6 e 7) permeiam mecanismos moleculares mais complexos como a perda de íntrons e transcrição reversa. Vários genes de baculovírus parecem ter sido adquiridos também de bactérias e plantas. Ambos os organismos estão, de alguma forma associados a insetos causando doenças, presentes na microbiota do trato digestório, ou como alimento (de Castro Oliveira et al., 2008). O mecanismo de aquisição de genes nesse contexto é completamente desconhecido.

Outra pergunta chave para o entendimento da aquisição gênica por virus está relacionada aos mecanismos de fixação do gene ao longo da evolução. Quais vantagens são conferidas por estas novas aquisições e como medi-las? Interessantemente, no decorrer destas linhas, dois genes *dut-tmk* (Capítulo 8) e *serpin* (Capítulo 5) foram encontrados como sendo capazes de modificar a infecção de um baculovírus prototípico. Neste contexto, características da infecção como número de vírus produzidos, nível de expressão de genes virais, virulência, e replicação foram avaliadas, e concluiu-se que a presença de tais genes, mesmo que num contexto não-natural, foi capaz de alterar o desempenho do vírus recombinante. Durante a evolução, quaisquer características que de alguma forma beneficiam a replicação viral e interfiram na manipulação do hospedeiro são positivamente selecionadas e fixadas. Por exemplo, a proteína inibidora de serino proteases foi capaz de controlar a resposta imune do inseto hospedeiro, inibindo a cascata de melanização que opsoniza antígenos presentes na hemolinfa do inseto, provavelmente protegendo o vírus de eventuais ataques (Ardisson-Araujo et al., 2015). Por outro lado, foi encontrado o primeiro baculovírus codificando para um transdutor de sinais, um receptor acoplado a proteína G (*disa038*, Capítulo 6) que provavelmente interfere na percepção da célula infectada ao ambiente.

Além disso, definições previamente estabelecidas podem mudar conforme novos genomas são sequenciados. Por exemplo, uma regra básica quanto à presença da glicoproteína de fusão GP64 (discutida nos Capítulos 6, 7 e 10) em baculovírus é a de que a proteína está presente apenas em alphabaculovírus do grupo I. Entretanto, neste trabalho, encontramos uma exceção a esta regra, que certamente redefinirá os conceitos para baculovírus: um homólogo funcional de *gp64* foi encontrado no genoma do betabaculovírus DisaGV. Betabaculovírus são definidos como não contendo GP64 como a glicoproteína de envelope principal (Rohrmann, 2013). Não está claro papel da GP64 na patologia da broca da cana de açúcar, uma vez que DisaGV também retém a proteína F, mas demonstramos que, apenas de com menor eficiência, a GP64 de DisaGV é funcional.

Não apenas a história evolutiva do vírus, como também da interação do homem com os insetos benéficos e pragas pode ser inferida pela genômica de baculovírus. Por exemplo, neste trabalho duas espécies de baculovírus já descritas em outros locais do mundo foram isoladas no Brasil e seus genomas sequenciados. Por reconstrução filogenética, encontramos que ambos Bombyx mori NPV (BmNPV) e Helicoverpa zea SNPV (HzSNPV) foram introduzidos no Brasil muito provavelmente por ação antrôpica. BmNPV é infectivo para o bicho da seda, *Bombyx mori* e causa intensas perdas na sericultura nacional e global. A cultura foi introduzida no Brasil por imigrantes japoneses e interessantemente, o isolado BmNPV-Brazilian é mais proximamente relacionado ao isolado japonês T3. Por outro lado, HzSNPV infecta lagartas de diferentes espécies polífagas do gênero *Helicoverpa*, que causa intensas perdas na agricultura. Este vírus foi isolado durante o primeiro surto de *Helicoverpa armigera* no país. Ambos os vírus BmNPV e HzSNPV foram provavelmente introduzidos no Brasil

junto com o hospedeiro inseto durante infecção não-letal, característica já descrita previamente para baculovírus.

Poucos trabalhos investigam o conteúdo gênico e o relacionam com a taxonomia do hospedeiro em nível de família. Especialização de patógenos aos seus hospedeiros pode ser consequência de co-evolução em longo termo, que é definida como uma evolução recíproca em espécies que interagem, dirigida por seleção natural (Herniou et al., 2004). Estas especializações podem ser refletidas na composição gênica. Baculovírus claramente co-evoluiu com o inseto hospedeiro em nível taxonômico de ordem; entretanto pouco se é investigado dessa co-evolução em nível de família. Neste trabalho, descrevemos pela primeira vez características genômicas associadas a um grupo específico de betabaculovírus infectivo para a família Noctuidae de lepidópteros (Capítulo 9). Uma clara expansão gênica aconteceu nesta família, levando ao surgimento dos maiores genomas entre os baculovírus. O controle fino da interação do vírus com o hospedeiro relativo ao limiar entre letalidade e latência parece ser mais complexo do que simplesmente replicar e causar a morte do hospedeiro. Este grupo de betabaculovírus apresenta baixa letalidade e longo tempo para causar a morte, restringindo a infecção ao tecido adiposo (Goldberg et al., 2002).

Em conclusão, a genômica e o estudo molecular básico de baculovírus têm influenciado também a compreensão de doenças associadas a humanos como câncer e infecções arbovirais. Por exemplo, baculovírus codificam em seu genoma uma série de proteínas inibidoras da resposta suicida, que bloqueiam direta ou indiretamente a resposta antiviral celular, os inibidores de apoptose (IAP) que foram descritos pela primeira vez em baculovírus (Crook et al., 1993) e estão associadas a várias neoplasias humanas

(Clem, 2015). Por outro lado, uma vez que baculovírus são vírus oralmente infectivos para insetos, a compreensão da rota de infecção viral pode ser estendida a outros vírus de inseto, como arbovírus causadores de doenças humanas uma vez que a mesma barreira de proteção inata de lepidóptera a ser transposta por baculovírus está conservada em vetores de doenças virais humanas, como mosquitos.

REFERÊNCIAS

Ardisson-Araujo, D. M., et al., 2015. Functional characterization of hesp018, a baculovirus-encoded serpin gene. J Gen Virol. 96**,** 1150-60.

Beperet, I., et al., 2014. Superinfection exclusion in alphabaculovirus infections is concomitant with actin reorganization. J Virol. 88**,** 3548-56.

Clem, R. J., 2015. Viral IAPs, then and now. Semin Cell Dev Biol. 39**,** 72-9.

Crook, N. E., et al., 1993. An apoptosis-inhibiting baculovirus gene with a zinc finger-like motif. J Virol. 67**,** 2168-74.

Cuartas, P. E., et al., 2015. The complete sequence of the first Spodoptera frugiperda Betabaculovirus genome: a natural multiple recombinant virus. Viruses. 7**,** 394-421.

de Castro Oliveira, J. V., et al., 2008. Structural and phylogenetic relationship of ORF 31 from the Anticarsia gemmatalis MNPV to poly (ADP-ribose) polymerases (PARP). Virus Genes. 37**,** 177-84.

Goldberg, A. V., et al., 2002. Effects of the Epap granulovirus on its host, Epinotia aporema (Lepidoptera: Tortricidae). J Invertebr Pathol. 80**,** 148-59.

Herniou, E. A., et al., 2001. Use of whole genome sequence data to infer baculovirus phylogeny. J Virol. 75**,** 8117-26.

Herniou, E. A., et al., 2003. The genome sequence and evolution of baculoviruses. Annu Rev Entomol. 48**,** 211-34.

Herniou, E. A., et al., 2004. Ancient coevolution of baculoviruses and their insect hosts. J Virol. 78**,** 3244-51.

Kamita, S. G., et al., 2005. A baculovirus-encoded protein tyrosine phosphatase gene induces enhanced locomotory activity in a lepidopteran host. Proc Natl Acad Sci U S A. 102**,** 2584-9.

Rohrmann, G. F., 2013. Baculovirus Molecular Biology. Bethesda (MD): National Center for Biotechnology Information (US). Available from: http://www.ncbi.nlm.nih.gov/books/NBK49500/.

Theze, J., et al., 2015. Gene acquisition convergence between entomopoxviruses and baculoviruses. Viruses. 7**,** 1960-74.

Westenberg, M., et al., 2007. Baculovirus envelope fusion proteins F and GP64 exploit distinct receptors to gain entry into cultured insect cells. J Gen Virol. 88**,** 3302-6.

Xu, P., et al., 2014. Densovirus Is a Mutualistic Symbiont of a Global Crop Pest (Helicoverpa armigera) and Protects against a Baculovirus and Bt Biopesticide. PLoS Pathog. 10**,** e1004490.

Anexo

- **Artigos**

Ardisson-Araujo, D. M., Melo, F. L., de Souza Andrade, M., Brancalhao, R. M., Bao, S. N. & Ribeiro, B. M. (2014b). Complete genome sequence of the first non-Asian isolate of Bombyx mori nucleopolyhedrovirus. *Virus Genes* **49**, 477-484. (CAPÍTULO 2)

Ardisson-Araujo, D. M., de Melo, F. L., Andrade M de, S., Sihler, W., Bao, S. N., Ribeiro, B. M. & de Souza, M. L. (2014a). Genome sequence of Erinnyis ello granulovirus (ErelGV), a natural cassava hornworm pesticide and the first sequenced sphingid-infecting betabaculovirus. *BMC Genomics* **15**, 856. (CAPÍTULO 3)

Ardisson-Araújo, D. M., Sosa-Gomez, D. R., Melo, F. L., Báo, S. N. & Ribeiro, B. M. (2015). Characterization of Helicoverpa zea single nucleopolyhedrovirus isolated in Brazil during the first old world bollworm (Noctuidae: *Helicoverpa armigera*) nationwide outbreak. *Virus Reviews & Research* **20**, 4. (CAPÍTULO 4)

Ardisson-Araujo, D. M., Rohrmann, G. F., Ribeiro, B. M. & Clem, R. J. (2015). Functional characterization of hesp018, a baculovirus-encoded serpin gene. *J Gen Virol* **96**, 1150-1160. (CAPÍTULO 5)

Ardisson-Araujo, D. M., Morgado Fda, S., Schwartz, E. F., Corzo, G. & Ribeiro, B. M. (2013a). A new theraphosid spider toxin causes early insect cell death by necrosis when expressed in vitro during recombinant baculovirus infection. *PLoS One* **8**, e84404.

Ardisson-Araujo, D. M., Rocha, J. R., da Costa, M. H., Bocca, A. L., Dusi, A. N., de Oliveira Resende, R. & Ribeiro, B. M. (2013b). A baculovirus-mediated strategy for full-length plant virus coat protein expression and purification. *Virol J* **10**, 262.

Braconi, C. T., Ardisson-Araujo, D. M., Paes Leme, A. F., Oliveira, J. V., Pauletti, B. A., Garcia-Maruniak, A., Ribeiro, B. M., Maruniak, J. E. & Zanotto, P. M. (2014). Proteomic analyses of baculovirus Anticarsia gemmatalis multiple nucleopolyhedrovirus budded and occluded virus. *J Gen Virol* **95**, 980-989.

de Oliveira, V. C., da Silva Morgado, F., Ardisson-Araujo, D. M., Resende, R. O. & Ribeiro, B. M. (2015). The silencing suppressor (NSs) protein of the plant virus Tomato spotted wilt virus enhances heterologous protein expression and baculovirus pathogenicity in cells and lepidopteran insects. *Arch Virol*.

Silva, L. A., Ardisson-Araujo, D. M., Tinoco, R. S., Fernandes, O. A., Melo, F. L. & Ribeiro, B. M. (2015). Complete genome sequence and structural characterization of a novel iflavirus isolated from Opsiphanes invirae (Lepidoptera: Nymphalidae). *J Invertebr Pathol* **130**, 136-140.

Teixeira Correa, R. F., Ardisson-Araujo, D. M., Monnerat, R. G. & Ribeiro, B. M. (2012). Cytotoxicity analysis of three Bacillus thuringiensis subsp. israelensis delta-endotoxins towards insect and mammalian cells. *PLoS One* **7**, e46121.

- **Patente:**

Domingues, R. A. S.; Ardisson-Araújo, Daniel Mendes Pereira; Santos, F. B.; Ribeiro, B. M.; Nagata, T. Peptídeos imunogênicos da proteína ns1 dos vírus da dengue, seu método de obtenção em baculovírus por meio da fusão com a poliedrina e seu uso para finalidade diagnóstica. 2014, Brasil. Patente: Privilégio de Inovação. Número do registro: BR10201401860, data de depósito: 30/07/2014. Instituição de registro: INPI - Instituto Nacional da Propriedade Industrial.

- **Capítulo de livro:**

Ribeiro, B. M.; Morgado, Fabrício da Silva; Ardisson-Araújo, Daniel Mendes Pereira; Silva, Leonardo A.; Cruz, F. S. P.; Chaves, L. S. C.; Quirino, M. S.; Andrade, M. S.; Corrêa, R. F. T.. Baculovírus para expressão de proteínas recombinantes em célula de inseto. In: Rodrigo R. Resende. (Org.). Biotecnologia Aplicada à Saúde. 1ed.São Paulo: Blucher, 2015, v. 2, p. 252-306.