ANDERSON FERNANDES DE BRITO

ANÁLISE DA DIVERSIDADE GENÔMICA DE ISOLADOS GEOGRÁFICOS DO NUCLEOPOLIEDROVÍRUS DE

Anticarsia gemmatalis (AgMNPV)

Dissertação apresentada ao Departamento de Microbiologia do Instituto de Ciências Biomédicas da Universidade de São Paulo, para obtenção do Título de Mestre em Ciências.

São Paulo 2013

ANDERSON FERNANDES DE BRITO

ANÁLISE DA DIVERSIDADE GENÔMICA DE ISOLADOS GEOGRÁFICOS DO NUCLEOPOLIEDROVÍRUS DE

Anticarsia gemmatalis (AgMNPV)

Dissertação apresentada ao Departamento de Microbiologia do Instituto de Ciências Biomédicas da Universidade de São Paulo, para obtenção do Título de Mestre em Ciências.

Área de concentração: Microbiologia

Orientador: Prof. Dr. Paolo Marinho de Andrade Zanotto

Versão corrigida. A versão original eletrônica encontra-se disponível tanto na Biblioteca do ICB quanto na Biblioteca Digital de Teses e Dissertações da USP (BDTD).

São Paulo 2013 DADOS DE CATALOGAÇÃO NA PUBLICAÇÃO (CIP) Serviço de Biblioteca e Informação Biomédica do Instituto de Ciências Biomédicas da Universidade de São Paulo

reprodução não autorizada pelo autor

Brito, Anderson Fernandes de.

Análise da diversidade genômica de isolados geográficos do nucleopoliedrovírus de *Anticarsia gemmatalis* (AgMNPV) / Anderson Fernandes de Brito. -- São Paulo, 2013.

Orientador: Prof. Dr. Paolo Marinho Andrade Zanotto.

Dissertação (Mestrado) – Universidade de São Paulo. Instituto de Ciências Biomédicas. Departamento de Microbiologia. Área de concentração: Microbiologia. Linha de pesquisa: Evolução molecular de Baculovírus.

Versão do título para o inglês: Genomic diversity analysis of geographic isolates of nucleopolyhedroviruses *Anticarsia gemmatalis* (AgMNPV).

1. Genômica 2. Evolução molecular 3. *Baculoviridae* 4. Diversidade genética 5. Vírus dos insetos 6. Sequenciamento genético I. Zanotto, Prof. Dr. Paolo Marinho Andrade II. Universidade de São Paulo. Instituto de Ciências Biomédicas. Programa de Pós-Graduação em Microbiologia III. Título.

ICB/SBIB0106/2013

UNIVERSIDADE DE SÃO PAULO INSTITUTO DE CIÊNCIAS BIOMÉDICAS

Candidato(a):	Anderson Fernandes de Brito.
Título da Dissertaç	ão: Análise da diversidade genômica de isolados geográficos do nucleopoliedrovírus de <i>Anticarsia gemmatalis</i> (AgMNPV).
Orientador(a):	Prof. Dr. Paolo Marinho Andrade Zanotto.
A Comissão Ju em sessão pú ()/	Igadora dos trabalhos de Defesa da Dissertação de Mestrado, blica realizada a///, considerou Aprovado(a) () Reprovado(a)
Examinador(a):	Assinatura: Nome: Instituição:
Examinador(a):	Assinatura: Nome: Instituição:
Presidente:	Assinatura: Nome: Instituição:



UNIVERSIDADE DE SÃO PAULO INSTITUTO DE CIÊNCIAS BIOMÉDICAS

Cidade Universitária "Armando de Salles Oliveira" Av. Prof. Lineu Prestes, 2415 – CEP. 05508-000 São Paulo, SP – Brasil Telefone :(55) (11) 3091-7733 - telefax : (55) (11) 3091-8405 e-mail: <u>ccp@icb.usp.br</u>

Comissão de Ética em Pe

CERTIFICADO DE ISENÇÃO

Certificamos que o Protocolo CEP-ICB N° 459/11 referente ao projeto intitulado: "*Metagenômica do nucleopoliedrovírus de Anticarsia gemmatalis (AgMNPV)*" sob a responsabilidade de Anderson Fernandes de Brito, foi analisado na presente data pela CEUA -COMISSÃO DE ÉTICA NO USO DE ANIMAIS e pela CEPSH- COMISSÃO DE ÉTICA EM PESQUISA COM SERES HUMANOS, tendo sido deliberado que o referido projeto não utilizará animais que estejam sob a égide da lei 11.794 de 8 de outubro de 2008, nem envolverá procedimentos regulados pela Resolução CONEP nº196 de 1996.

São Paulo, 17 de maio de 2011.

PROF. DR. WOTHAN TAVARES DE LIMA Coordenador da CEUA - ICB/USP

PROF. DR. PAOLO M.A ZANOTTO Coordenador da CEPsh - ICB/USP

À minha mãe (Marlene Fernandes) e ao meu pai (Lírio Brito): por serem exemplos de caráter, dedicação, e perserverança.

À minha irmã (Andréia) e a todos os meus familiares: pela compreensão, confiança e incentivos incondicionais, sem os quais eu jamais poderia alcançar os meus objetivos.

AGRADECIMENTOS

A toda a minha família e amigos de Brasília, que mesmo tão longe fisicamente, fizeram-se intensamente presentes durante minhas curtas visitas à minha cidade natal. Sem o apoio, compreensão, e companheirismo de vocês, minhas atuais conquistas jamais seriam possíveis.

Ao meu orientador de mestrado, Paolo Zanotto, por ter prontamente confiado no meu potencial; por ter dado o suporte indispensável à execução do meu trabalho; e por servir de incentivo para que eu superasse meus limites pessoais e acadêmicos.

A todos os atuais e ex-membros do LEMB com os quais interagi em menor ou maior escala ao longo destes mais de dois anos. À Carla, por toda ajuda dada, não só nas etapas experimentais do meu trabalho, como pelo apoio pessoal e financeiro a mim oferecido durante os primeiros meses de minha estadia em São Paulo. Ao Julian, pelas trocas de experiências e por todo apoio prestado ao longo do mestrado: admiro sua postura como profissional e cientista. À Dany e Victor, por todo apoio e atenção prestada dentro e fora do laboratório, seja para assuntos acadêmicos como para conversas alegres e descontraídas. Ao Atila e Caio, pelas sugestões e auxílios prestados no decorrer do meu projeto.

Aos meus orientadores de iniciação científica Marlinda Lobo e William Sihler, por terem tão pacientemente me introduzido à Virologia e ao ramo de estudo com Baculovírus.

Aos funcionários do ICB: pessoal da portaria/segurança; limpeza/manutenção; e secretarias. Meus sinceros agradecimentos a todos, que dentro de suas respectivas funções, possibilitaram um ambiente de trabalho funcional e agradável. Em âmbito laboratorial, agradeço especialmente ao Dino pela dedicação nos trabalhos técnicos, essenciais à agilidade das atividades experimentais.

Aos Doutores Meik Dilcher e Manfred Weidmann, da Universidade of Göttingen (Alemanha), pelo pirosequenciamento das amostras, e por terem sido tão solícitos ao prestar suporte antes e depois do sequenciamento.

Ao Dr. Daniel Ricardo Sosa-Gómez, da Embrapa Soja, pela disponibilização das amostras.

A todos os meus amigos da Escola de Dança Locus, por terem me apresentado a magia da dança, assim como pela amizade e carinho oferecidos nestes últimos anos.

A todos os meus colegas da república LEBAS, pela companhia e por oferecerem um ambiente enriquecedor, que me possibilitou o contato com diferentes formas de viver e maneiras de pensar.

À CAPES e à FAPESP pelo apoio financeiro.

"Eu sei que a gente se acostuma. Mas não devia.

A gente se acostuma a coisas demais, para não sofrer. Em doses pequenas, tentando não perceber, vai afastando uma dor aqui, um ressentimento ali, uma revolta acolá. A gente se acostuma para não se ralar na aspereza, para preservar a pele. Se acostuma para evitar feridas, sangramentos, para poupar o peito. A gente se acostuma para poupar a vida. Que aos poucos se gasta, e que, gasta de tanto costumar, se perde de si mesma" Marina Colasanti

RESUMO

BRITO, A. F. Análise da diversidade genômica de isolados geográficos do nucleopoliedrovírus de *Anticarsia gemmatalis* (AgMNPV). 2013. 166 f. Dissertação (Mestrado em Microbiologia) - Instituto de Ciências Biomédicas, Universidade de São Paulo, São Paulo, 2013.

O Anticarsia gemmatalis multiple nucleopolyhedrovirus é o bioinseticida viral mais amplamente utilizado no mundo que tem sido empregado no Brasil por mais de 30 anos como agente de controle biológico da lagarta-da-soja (Anticarsia gemmatalis). Diversos variantes derivados de populações naturais deste vírus já foram analisados previamente por meio de perfis de restrição de DNA, evidenciando a presença de diversos polimorfismos, muitos dos quais, no entanto, não foram caracterizadas em detalhes. Um destes variantes, o protótipo da espécie (AgMNPV-2D), teve seu genoma sequenciado em 2006. O presente trabalho de mestrado avaliou a diversidade genômica de 17 isolados geográficos selvagens de AgMNPV obtidos entre os anos de 1980 a 1990, no Brasil, Argentina e Uruguai. As amostras coletadas são populações virais genéticamente heterogêneas de ocorrência natural em cultivos de soja sem histórico de uso de baculovírus como bioinseticida. Entre os 17 genomas reconstruídos a alta compactação da informação genética manteve-se conservada ao longo da história evolutiva de AgMNPV, com a maioria das ORFs dispostas de maneira contígua ou sobrepostas umas as outras. Rearranjos estruturais significativos, tais como inversões e translocações, não foram observados. Por outro lado, deleções e inserções de uma ou mais bases foram comuns, muitas das quais mudaram o quadro de leitura de algumas ORFs. A partir de dados do proteoma estrutural do isolado 2D, foi possível identificar um novo peptídeo codificado por uma ORF aninhada e conservada em todos os isolados analizados (ORF29b). Um novo gene hipotético com possível origem em lepidóptero (ORF25b) foi identificado em dois genótipos; e o gene bro-a mostrou-se ausente em muitos isolados. Análises de diversidade genética interpopulacionais mostraram que a maioria dos genes de AgMNPV é pouco variável, sendo os genes mais diversos localizados em loci que coincidem com regiões variáveis mapeadas em estudos anteriores. Os resultados obtidos reforçam a percepção de que grandes vírus de DNA evoluem principalmente sob baixas taxas de substituição, e por meio de duplicações, ganhos e perdas gênicas.

Palavras-chave: AgMNPV. *Anticarsia gemmatalis*. Baculovírus. Diversidade Genética. Evolução Molecular. Genoma.

ABSTRACT

BRITO, A. F. Genomic diversity analysis of geographic isolates of nucleopolyhedroviruses *Anticarsia gemmatalis* (AgMNPV). 2013. 166 p. Masters thesis (Microbiology) - Instituto de Ciências Biomédicas, Universidade de São Paulo, São Paulo, 2013.

The Anticarsia gemmatalis multiple nucleopolyhedrovirus is the most widely used viral bioinseticide in the world. In Brazil, this virus has been used during the last 30 years as a biological control agent against velvetbean caterpillar (Anticarsia gemmatalis). Several variants derived from natural viral populations were previously analyzed by DNA restriction analysis, which evidenced the presence of multiple polymorphisms; many of which were not characterized in details. One of these variants, the prototype of AgMNPV (isolate 2D), was sequenced in 2006. This Masters degree project evaluated the genomic diversity of 17 AgMNPV geographic isolates obtained from 1980 to 1990, in the Brazil, Argentina and Uruguay. The samples are essentially heterogeneous viral populations of natural occurrence in soybean crops with no history of use of baculovirus as bioinseticide. Among 17 genomes reconstructed the high compaction of the genetic information was maintained during AgMNPV evolutionary history, with most ORFs are arranged contiguously or overlapping each other. Significant structural rearrangements, such as inversions and translocations, were not observed. Otherwise, deletions and insertions of one or more bases were common, many of which changed the reading frame of some ORFs. Using available data from a strutuctral proteome study, it was possible to identify a novel peptide codified by a nested ORF conserved in all analyzed isolates (ORF29b). A new hypothetical gene with possible lepidopteran origin (ORF25b) was identified in two genotypes; and the gene *bro-a* is absent in many isolates. Interpopulational genetic diversity analysis pointed that most AgMNPV genes shows little variation, and the highly diverse genes are located in *loci* which coincide with variable regions mapped in previous studies. The results obtained reinforce the perception that large DNA viruses evolve mainly under low rates of substitution, and through duplications, gain and loss of genic regions.

Keywords: AgMNPV. *Anticarsia gemmatalis*. Baculovirus. Genetic Diversity. Genome. Molecular Evolution.

SUMÁRIO

1 INTRODUÇÃO	14
1.1 VISÃO GERAL SOBRE A FAMÍLIA BACULOVIRIDAE	14
1.1.1 Morfologia e classificação dos baculovírus	14
1.1.2 Características dos genomas dos baculovírus	16
1.1.3 Composição gênica	17
1.2 CICLO BIOLÓGICO E ECOLOGIA DOS BACULOVÍRUS	20
1.3 REGULAÇÃO DA EXPRESSÃO GÊNICA	22
1.4 EVOLUÇÃO DOS BACULOVÍRUS	22
1.5 BACULOVÍRUS Anticarsia gemmatalis MNPV	25
2 OBJETIVOS	29
3 MATERIAIS E MÉTODOS	30
3.1 AMOSTRAS VIRAIS	30
3.2 AMPLIFICAÇÃO DOS ESTOQUES VIRAIS IN VIVO	31
3.3 PURIFICAÇÃO DOS CORPOS DE OCLUSÃO	33
3.4 EXTRAÇÃO DE DNA DAS AMOSTRAS VIRAIS	34
3.5 PIROSEQUENCIAMENTO	34
3.6 REMOÇÃO DOS ADAPTADORES	36
3.7 MAPEAMENTO DAS LEITURAS SOBRE O GENOMA REFERÊNCIA	37
3.8 RECONSTRUÇÃO DOS GENOMAS PREDOMINANTES	41
3.8.1 Montagens preliminares	41
3.8.2 Setorização do genoma de AgMNPV-2D	42
3.8.3 Montagem dos genomas	43
3.8.4 Junção de contigs dos setores genômicos	45
3.9 CORREÇÃO DE INSERÇÕES E DELEÇÕES EM REGIÕES CODIFICANTES	46
3.10 ANOTAÇÕES GENÔMICAS	46
3.11 IDENTIFICAÇÃO DE NOVAS ORFS POR ESPECTROMETRIA DE MASSA.	46
3.12 ANÁLISE DE DOMÍNIOS PROTÉICOS	47
3.13 ESTIMATIVAS DE DIVERSIDADE GENÉTICA	47
3.14 ORGANIZAÇÃO GENÔMICA E SINTENIA	48
4 RESULTADOS	49
4.1 DADOS DO PIROSEQUENCIAMENTO	49
4.1.1 Fragmentação do DNA genômico	49

4.1.2 Características do conjunto de dados	49
4.2 LEITURAS NÃO CORRESPONDENTES A AGMNPV-2D	52
4.3 ORGANIZAÇÃO GENÔMICA DOS ALFABACULOVÍRUS DO GRUPO I	53
4.3.1 Sintenia dos genomas	53
4.3.2 Regiões de transcrição sobreposta (ROTs)	54
4.4 ORGANIZAÇÃO GENÔMICA DOS ISOLADOS DE AgMNPV	57
4.4.1 Colinearidade genômica de AgMNPV	57
4.4.2 Características gerais dos genomas de AgMNPV	
4.5 REGIÕES CODIFICANTES DOS ISOLADOS DE AgMNPV	71
4.5.1 Sobreposição de regiões codificantes e as ROTs	
4.5.2 Conjunto de genes presentes entre os isolados de AgMNPV	
4.6 ORFs DE SUPOSTOS NOVOS GENES DE AgMNPV	
4.7 VARIAÇÕES ESTRUTURAIS EM RELAÇÃO AO AgMNPV-2D	
4.7.1 A ORF2b de AgMNPV-2D	
4.7.2 Alterações nos loci de bro-a, bro-b e bro-c	82
4.7.3 Separação da ORF21	
4.7.4 Truncamento do gene hisP	87
4.7.5 Alterações na ORF64	88
4.7.6 Fusão das ORFs 101 e 102 de he65	90
4.7.7 Polimorfismos estruturais de pe-38	
4.8 EVIDÊNCIA DE ORFs ANINHADAS EM BACULOVÍRUS	
4.9 Ag25b, UM GENE ÚNICO DE AgMNPV	
4.10 DIVERSIDADE GENÉTICA EM AgMNPV	
4.10.1 Diversidade em genes core, não-core, e supostos genes	
4.10.2 Diversidade por categorias funcionais	101
5 DISCUSSÃO	105
5.1 O SEQUENCIAMNETO DE GENOMAS VIRAIS EM LARGA ESCALA	105
5.1.1 Montagens genômicas num contexto de desequilíbrio de ligação	105
5.2 ORGANIZAÇÃO GENÔMICA E COMPOSIÇÃO GÊNICA DE AgMNPV	107
5.2.1 A conservação da ordem gênica em AgMNPV	107
5.2.2 A compactação da informação genética e as ROTs	108
5.2.3 Diferentes níveis de organização das regiões codificantes	111
5.3 OS PADRÕES EVOLUTIVOS DE AgMNPV	112

5.3.1 Origens e implicações evolutivas das sobreposições entre genes	112
5.3.2 Fusões e fissões gênicas	116
5.3.3 Perda e ganho de fragmentos genômicos em pe-38	119
5.3.4 Variações haplotípicas envolvendo do gene bro-a	120
5.3.5 Um possível novo gene codificante de RING Finger	123
5.3.6 Alterações em hotspots de recombinação homóloga	125
5.4 A VARIABILIDADE GENÉTICA EM AgMNPV	129
6 CONCLUSÕES	132
REFERÊNCIAS*	134
ANEXO A – Tabelas Suplementares	151
ANEXO B – Artigo	155

1 INTRODUÇÃO

1.1 VISÃO GERAL SOBRE A FAMÍLIA BACULOVIRIDAE

1.1.1 Morfologia e classificação dos baculovírus

Os membros da família *Baculoviridae* infectam artrópodes, em especial insetos das ordens *Lepidoptera*, *Hymenoptera* e *Díptera*. Diversos baculovírus foram isolados de mais de 600 espécies de insetos destas ordens (MARTIGNONI & IWAI, 1986). Vírus com características semelhantes à baculovírus já foram observados em outras ordens de organismos da classe *Insecta*, como: *Coleoptera*, *Neuroptera*, *Siphonaptera*, *Thysanura* e *Trichoptera*. Muitos dos vírus que infectam estas ordens ainda não foram caracterizados a nível molecular, logo não é possível afirmar que são de fato baculovírus (JEHLE et al., 2006).

Os capsídeos dos baculovírus medem de 30 a 60 nm de diâmetro e de 250 a 300 nm de comprimento, possuem forma de bastão, morfologia da qual deriva o nome da família (*baculo*, do latim *baculum* = bastão) (BILIMORIA, 1991; BONNING, 2005; HARRAP, 1972; JEHLE et al., 2006). A maioria dos baculovírus apresenta dois fenótipos de partículas virais geneticamente idênticas, porém morfologicamente distintas: BV (*Budded Virus*), partículas envelopadas compostas por um único nucleocapsídeo por envelope, que são responsáveis pela infecção célula-célula; e ODV (*Occlusion-Derived Virus*), que são partículas envelopadas com um ou vários nucleocapsídeos por envelope. Estas são adicionalmente envoltas por uma matriz protéica cristalina que forma o corpo de oclusão (*Occlusion Body* - OB). Os OBs participam da infecção hospedeiro-hospedeiro, geralmente medem de 1 a 5 µm de diâmetro, e sua matriz paracristalina garante estabilidade física e biológica às partículas virais (BILIMORIA, 1991; O'REILLY *et al.*, 1993).

Até recentemente, a família *Baculoviridae* era subdividida em dois grupos: *Nucleopolyhedrovirus* (NPV) e *Granulovirus* (GV). Os NPVs por sua vez eram classificados como SNPV (S = single, apenas um nucleocapsídeo) ou MNPV (M = *Multiple*, múltiplos nucleocapsídeos), em referência a quantidade de nucleocapsídeos contidos nas partículas ODV (Figura 1 e 2). Atualmente, *Baculoviridae* está organizada em quatro gêneros: *Alphabaculovirus* (NPVs de Lepidoptera), *Betabaculovirus* (GVs de Lepidoptera), *Gammabaculovirus* (NPVs de Hymenoptera) e *Deltabaculovirus* (NPVs de Diptera) (JEHLE et al., 2006; ROHRMANN, 2011).



Figura 1 - Eletromicrografia de um corpo de oclusão (OB) de nucleopoliedrovírus.

As setas indicam nucleocapsídeos imersos na matriz protéica de poliedrina de um OB. Fonte: pt.wikipedia.org (Domínio público).



Figura 2 – Partículas virais de nucleopoliedrovírus (MNPV).

Em (1) observa-se o aspecto global da morfologia dos poliedros (corpos de oclusão - OB), representado em corte transversal e ampliado em (2). Internamente estão representadas as partículas ODV (*Occlusion-Derived Virus*), em cortes transversais e longitudinais, contendo variadas quantidades de nucleocapsídeos. Em (3) observa-se corte longitudinal e ampliado de uma partícula viral ODV, contendo nucleocapsídeos que encerram diferentes genomas, gerados em decorrência de recombinações entre diferentes variantes (em cores). Estas partículas, liberadas pela dissolução dos poliedros, iniciam o estágio de infecção primária, e em seguida podem dar origem às partículas do fenótipo BV (*Budded Virus*). Os BVs (4) procederão as infecções secundárias, célula a célula, promovendo a síntese adicional de novos poliedros.

1.1.2 Características dos genomas dos baculovírus

Os baculovírus apresentam genoma composto por DNA dupla fita, circular, superenovelado, com tamanho variando em torno de 80 a 180 Kbp, o qual é capaz de codificar de 90 a 183 genes, organizados de maneira contígua, podendo inclusive apresentarem-se sobrepostos uns aos outros (MIELE et al., 2011; ROHRMANN, 2011; VOLKMAN et al., 1995). O conteúdo GC dos genomas de baculovírus varia de 32,4% a 57,5% (HAYAKAWA *et al.*, 2000; VAN OERS & VLAK, 2007). Seus genes estão dispostos

em ambas as fitas de DNA, sem qualquer padrão de agrupamento relativo a promotores (precoces ou tardios) ou categorias funcionais, estando homogeneamente distribuídos nos sensos positivo e negativo dos genomas (COHEN *et al.*, 2009; VAN OERS & VLAK, 2007). Por convenção, a ORF (*Open Reading Frame*) do gene 1 dos baculovírus corresponde a ORF de *polh* (VLAK & SMITH, 1982), gene que codifica a proteína que majoritariamente compõe o corpo de oclusão viral (ROHRMANN, 1986).

No genoma de muitos baculovírus é comum a ocorrência de *loci* constituídos por múltiplas repetições de sequências em tandem, ricas em AT, e com motivos palindrômicos. Estes *loci*, por sua vez, apresentam-se em múltiplas cópias, espalhadas em várias posições dos genomas, caracterizando "regiões altamente repetitivas" (*homologous regions, hrs*) (COCHRAN & FAULKNER, 1983; POSSEE & ROHRMANN, 1997; RODEMS & FRIESEN, 1995). A quantidade de cópias presentes nos genomas de baculovírus varia de três (LANGE & JEHLE, 2003) a 17 (PANG et al., 2001), e juntas podem compreender de 3 a 7% do material genômico viral (HAYAKAWA *et al.*, 2000); por outro lado, em alguns vírus não foram encontrados *loci* semelhantes àquelas regiões altamente repetitivas (LUQUE *et al.*, 2001; VAN OERS *et al.*, 2005; WILLIS *et al.*, 2005; WORMLEATON *et al.*, 2003). Estas regiões têm sido apontadas como intensificadores da expressão gênica precoce, e como origens de replicação do genoma viral (POSSEE & ROHRMANN, 1997).

Excetuando-se as regiões não-codificantes identificadas como *hrs*, os espaços intergênicos podem corresponder de 4 a 17 Kbp do genoma dos baculovírus (HAYAKAWA *et al.*, 2000), locais onde estão principalmente localizados elementos de regulação da expressão gênica, tais como promotores e terminadores de transcrição (NELSON *et al.*, 2004).

1.1.3 Composição gênica

Os genes de baculovírus podem ser principalmente divididos em três grupos: (*i*) genes compartilhados por todos os baculovírus; (*ii*) genes presentes apenas em vírus filogeneticamente próximos ou relacionados (dos 4 gêneros da família); e (*iii*) genes restritos a uma única espécie (HAYAKAWA *et al.*, 2000; MIELE *et al.*, 2011). O viés de uso de códons é bastante variável entre genes de um mesmo vírus, e até mesmo entre genes homólogos de diferentes vírus (LEVIN & WHITTOME, 2000).

Entre os 58 genomas de baculovírus sequenciados e disponíveis no *GenBank*, foram identificados 37 genes compartilhados por todos os membros da família *Baculoviridae* (GARAVAGLIA et al., 2012). Estes desempenham funções essenciais a infectividade viral, sendo denominados *core genes* (BLISSARD & ROHRMANN, 1990; COHEN *et al.*, 2009; VAN OERS & VLAK, 2007).

Quantos às funções desempenhadas pelos produtos dos genes, estes podem ser classificados em cinco categorias funcionais, de acordo as suas principais funções (COHEN *et al.*, 2009; GARAVAGLIA *et al.*, 2012; MIELE *et al.*, 2011; VAN OERS & VLAK, 2007): Replicação de DNA; Transcrição; Estrutural; Modulação (do metabolismo do hospedeiro); e Auxiliar (Figura 3).



Α		50			
	11	Categorias	funcionais		
		Boplicação	Desceptoride	35	
		Auxiliar	Desconnecido		
	27	Auxiliar	Estrutural		
	21		Modulação		
		16	11		
_					
в					
003 鱼 lef2	007 📕 bro-b	024 🔳 lef11	082 兽 vp91/p95	002 🔳 Ag2	075 🔳 Ag75
019 单 lef1	011 📕 bro-c	025 📕 39K/pp31	084 📕 p15	010 🔳 Ag10	077 🔍 Ag77
042 📕 dbp	028 📕 hisP	039 📕 lef6	086 兽 vp39	014 🔳 Ag14	078 🔳 Ag78
046 📕 gta	030 📕 ctl-2	044 🔍 p47	091 😑 odve25	015 🔳 Ag15	081 🔳 tlp20
065 ONApol	031 📕 parp-like	045 📕 lef12	093 🔴 pif4	016 📕 Ag16	083 📕 Ag83
067 📕 lef3	032 sod	051 • lef8	094 9 38k	020 Ag20	088 🔳 Ag88
092 helicase	033 bro-d	054 📕 lef10	096 9 <i>p6.9</i>	021 Ag21	090 p 18
123 ef/	035 Ag35	062 e lef9	097 p 40/c42	02/ Ag2/	104 Ag104
129 aikexo	030 Ag30	009 met	100 a p8//vp80	037 Ag37	109 Ag109
148 ne-38	061 25kfp	070 • Vij1 087 • lef4	107 a odv-ec/13	038 Ags8	110 Ag110
148 pe-38	063 hro-e	095 e lef5	108 4 <i>a</i> 108	041 Ag41	111 Ag111
009 b tp1/bvp	066 desmoplakin	143 i e1	114 • pif3	048 Aq48	113 Aq113
013 arif1	085 c a30	145 📕 ie2	117 • pif1	050 🔳 etm	116 🔳 Aq116
018 📕 egt	099 🔵 p48	151 📕 pk1	124 📕 gp64	052 🔳 Ag52	118 🔳 Ag118
026 📕 ubi	101 📕 he65	001 📕 polh	125 p 24	056 🔳 Ag56	119 🔳 Ag119
029 📕 fgf	103 📕 pnk/pnl	012 🛑 pif2	127 📕 pp34/pep	057 🔳 Ag57	120 🔳 Ag120
034 📕 iap3	105 📕 bro-f	017 📕 odve26	128 📕 p25	058 🔳 Ag58	121 🔳 Ag121
040 📕 iap1	115 📕 bro-g	022 📕 F-protein	133 📕 p10	059 🔳 Ag59	130 🔳 Ag130
070 📕 iap2	122 📕 v-trex	049 📕 odve66	134 🛑 p74	060 📕 Ag60	131 📕 Ag131
089 🔵 p33	126 📕 gp16	053 🔴 Ag53	138 🛑 p49	064 🔳 Ag64	132 🔳 p26
098 📕 p12	135 bro-h	055 🛑 vp1054	139 🛑 odve18	071 📕 Ag71	142 📕 Ag142
004 Ag4	137 i e0	068 e pif6	140 • odve27	072 Ag72	146 🔳 Ag146
005 Ag5	141 Ag141	079 g p41	144 • odve56/pif5	073 Ag73	147 Ag147
006 🔳 bro-a	023 🔳 Ag23	080 🛡 Ag80	152 📕 1629capsid	074 🔳 Ag74	150 🔳 Ag150

Genes com mais de uma função tiveram sua categoria definida com base na sua principal função segundo a literatura. A) Proporção de genes entre as 5 categorias funcionais e com função desconhecida. Os números indicam a quantidade de genes em cada categoria. B) Lista de genes associados à AgMNPV. As cores correspondem a apresentada em A. A esquerda, o número da ORF em AgMNPV, a direita o nome do gene. Os 37 genes *core* estão identificados em **negrito** e por um círculo (\bullet) ao centro.

1.2 CICLO BIOLÓGICO E ECOLOGIA DOS BACULOVÍRUS

O início do ciclo biológico dos baculovírus pode ser definido pela ingestão de corpos de oclusão (OB) pelos hospedeiros, que em sua maioria são insetos em fase larval (juvenil). Nesta fase, o organismo está passando por constantes divisões e diferenciações celulares, condições ideais à replicação viral, nas quais os baculovírus evoluíram e se adaptaram. Depois de ingeridos junto com o alimento, os OB são dissolvidos assim que atingem o ambiente alcalino do intestino médio do hospedeiro, uma das poucas regiões que permitem o contato direto do vírus com as células do epitélio colunar, uma vez que a membrana peritrófica é porosa (FRIESEN & MILLER, 2001; ROHRMANN, 2011).

Com auxílio de metaloproteases virais, os vírions ODV liberados cruzam a membrana peritrófica, uma camada de consistência gelatinosa que reveste o intestino médio; ligam-se a receptores celulares; e se fundem com as microvilosidades das células do epitélio colunar intestinal (VOLKMAN, 2007). Quando os nucleocapsídeos alcançam o citosol, estes são direcionados ao núcleo, com o auxílio ativo de filamentos de actina e miosina do citoesqueleto, onde penetram por meio dos poros nucleares. O material genético viral é desnudado, possibilitando a transcrição, tradução e replicação do genoma viral neste sítio celular (GRANADOS & LAWLER, 1981).

Após a entrada nas células intestinais, o vírus leva de 30 minutos a duas horas para produzir sua primeira progênie de BVs, os quais após as infecção das células basais do intestino médio, irão brotar na hemocele, possibilitando a infecção sistêmica no corpo do inseto (FLIPSEN *et al.*, 1995; GRANADOS & LAWLER, 1981). Em seguida, outros tecidos e órgãos da larva são sequencialmente infectados: células traqueais; hemócitos (hemolinfa); músculos; tecido adiposo (corpo gorduroso); túbulos de Malpighi; e gânglios nervosos (SOARES & RIBEIRO, 2005). Nestes diferentes sítios de replicação viral são originadas partículas virais na forma oclusa. A infecção de um único inseto pode produzir por volta de 2 $\times 10^9$ poliedros, os quais mediante a desintegração do corpo dos insetos mortos servirão de inóculo para infecções futuras (Figura 4) (KUNIMI *et al.*, 1996).



Figura 4 – Ciclo biológico de um nucleopoliedrovírus.

Fonte: pt.wikipedia.org (com permissão).

Fatores bióticos e abióticos podem contribuir para a dispersão viral nos ecossistemas naturais (FUXA, 2004). Os corpos de oclusão podem ser ingeridos por pássaros ou outros insetos que se alimentam dos hospedeiros moribundos ou mortos, promovendo a dispersão viral num raio de poucos centímetros (em decorrência de insetos predadores) (LEE & FUXA, 2000) a até dezenas de quilômetros (transportados por pássaros) (ENTWISTLE et al., 1983).

Uma vez dispersos no ambiente, os poliedros podem persistir viáveis por 5 a 40 anos no solo (JAQUES, 1967; THOMPSON *et al.*, 1981). A chuva é um dos principais fatores responsáveis pelo transporte de inóculos virais do solo para a parte área de plantas como a soja (YOUNG & YEARIAN, 1986). Especialmente em cultivos desta planta, um estudo prévio apontou que corpos de oclusão presentes no solo por pelo menos 2 anos foram capazes de iniciar epizootias por AgMNPV em populações de lagarta-da-soja (*Anticarsia gemmatalis*) (FUXA & RICHTER, 1999).

1.3 REGULAÇÃO DA EXPRESSÃO GÊNICA

Em geral, a replicação dos baculovírus pode ser dividida em três fases: *early, late* e *very late*. A fase *early* (precoce) em média se estende de 0 a 6 h.p.i (horas pós infecção): é a etapa que precede a replicação do genoma viral (FRIESEN & MILLER, 1986;2001). Genes *early* são regulados pela ação transativadora de proteínas codificadas pelos genes *ie1, ie0, ie2* e *pe-38* (CARSON *et al.*, 1991; KRAPPA *et al.*, 1995). Os genes desta fase são transcritos pela RNA polimerase II celular, e seus produtos atuam em processos de: replicação do DNA viral; expressão de genes *late* e *very late;* e modificação do metabolismo celular, por meio de vias de sinalização (BLISSARD *et al.*, 1992; FRIESEN & MILLER, 1986;2001; HUH & WEAVER, 1990).

A fase *late* (tardia), que compreende o período de 6 a 24 h.p.i, é marcada pela síntese de DNA viral, assim como pela expressão de proteínas estruturais que irão compor os nucleocapsídeos dos novos vírus em formação (FRIESEN & MILLER, 1986; LU & MILLER, 1997). É nesta fase — onde os genes são transcritos por meio da uma RNA polimerase viral (GUARINO et al., 1998) — que são formadas a maioria das partículas BV, que brotam da membrana celular e espalham a infecção por diversas células e tecidos do hospedeiro (O'REILLY *et al.*, 1993). Na fase *very late* (muito tardia), que tem início após 24 h.p.i., são expressas proteínas que atuarão na formação dos corpos de oclusão viral, em especial a codificada pelo gene *polh*. Neste estágio da infecção o *polh* é super expresso, tendo como produto uma proteína que compõe majoritariamente os OB (granulina, em *betabaculovirus*, ou poliedrina, para os demais gêneros) (BLISSARD & ROHRMANN, 1990; FRIESEN & MILLER, 1986;2001; LU & MILLER, 1997; O'REILLY *et al.*, 1993).

1.4 EVOLUÇÃO DOS BACULOVÍRUS

Os baculovírus compreendem um dos grupos de vírus de insetos mais diversos (BLISSARD & ROHRMANN, 1990). Diferentemente de muitos vírus de RNA, vírus de DNA, em especial aqueles com grandes genomas, como os baculovírus, apresentam baixas taxas de mutação e substituição (DUFFY *et al.*, 2008). Comumente, a história evolutiva

destes vírus é delineada por frequentes eventos de transferência horizontal de genes (HUGHES & FRIEDMAN, 2003) e sequência não-codificantes (*transposons*) de origem celular (BLISSARD & ROHRMANN, 1990); duplicação e perda de genes e fragmentos gênicos (HAYAKAWA *et al.*, 2000); e recombinações entre variantes virais (CROIZIER & RIBEIRO, 1992; HAJOS *et al.*, 2000; KAMITA *et al.*, 2003).

Os genomas dos grandes vírus de DNA evoluem pela agregação contínua de genes celulares (HUGHES, 2002b; SHACKELTON & HOLMES, 2004; ZANOTTO & KRAKAUER, 2008). Para pelo menos seis genes compartilhados por muitos baculovírus foi possível inferir a homologia destes com genes celulares. Análises filogenéticas dos genes *DNA ligase; rr1 (ribonucleotide reductase 1)*; e *iap* mostraram que estes, respectivamente, não formam um grupo monofilético em filogenias com ortólogos em outros organismos, o que sugere eventos independentes de perdas e ganhos destes genes entre os membros de *Baculoviridae* (HUGHES, 2002a; HUGHES & FRIEDMAN, 2003; VAN STRIEN *et al.*, 1997). Em análises semelhantes, os genes *chitinase, gta*, e *egt* apresentam-se como clados monofiléticos, indicando que tais genes estiveram incorporados no genoma de ancestrais comum aos baculovírus (HUGHES & FRIEDMAN, 2003). Além destes, conjuntos de genes das mais variadas categorias funcionais foram adquiridos por diferentes linhagens de baculovírus, sendo muitos destes exclusivamente encontrados numa única espécie de vírus (HERNIOU et al., 2003).

Baculovírus possuem uma grande dinâmica de ganho e perda de genes, e muitos dos genes adquiridos podem desempenhar um importante papel na colonização de determinados grupos de hospedeiros (KRAKAUER & ZANOTTO, 2009; ZANOTTO & KRAKAUER, 2008). Um importante grupo de genes com interessante histórico de ganhos, perdas e duplicações são os componentes da família multigênica *bro* (*Baculovirus Repeated ORFs*) (IYER *et al.*, 2002; LOPEZ FERBER *et al.*, 2001). Os genes *bro* estão presentes em múltiplas cópias não só em baculovírus, como também em outras famílias virais, como *Poxviridae* (entomopoxvírus), *Iridoviridae* e *Phycodnaviridae*; assim como em bacteriófagos e elementos transponíveis bacterianos (das famílias IS3 e IS5). Entretanto, os eventos por trás dos ganhos e perdas de cópias destes genes ainda são incertos (BIDESHI *et al.*, 2003; IYER *et al.*, 2002). Além dos genes *bro*, uma série de outros genes estão implicados em eventos de ganhos, perdas ou duplicações, como *iap, ctl, enhancin, p10, fgf e ptp* (HAYAKAWA *et al.*, 2000). Muitos destes genes tem sido classificados como genes auxiliares, sendo, portanto, elementos que proporcionam vantagens em termos de infectividade, que no entanto não são essenciais a

replicação viral, fato que explicaria a perda destes genes ao longo da história evolutiva de algumas linhagens (HERNIOU et al., 2003).

Recombinações entre variantes de baculovírus são eventos bem documentados em situações de coinfecção (CROIZIER & RIBEIRO, 1992; HAJOS *et al.*, 2000; KAMITA *et al.*, 2003; KONDO & MAEDA, 1991; LEE & MILLER, 1979; MCINTOSH *et al.*, 1987; MERRYWEATHER-CLARKE *et al.*, 1994; MUNOZ *et al.*, 1997; SMITH & SUMMERS, 1980). Em condições ótimas *in vitro*, a frequência de geração de recombinantes pode chegar a 47%, correlacionando-se positivamente com a multiplicidade de infecção (MOI) utilizada (CROIZIER & RIBEIRO, 1992; HAJOS *et al.*, 2000; KAMITA *et al.*, 2003). Por outro lado, em condições sub-ótimas, como em infecções *in vivo*, as recombinações podem ser até 50 vezes menos frequentes (KAMITA *et al.*, 2003).

Eventos de recombinação e rearranjos estruturais em genomas de baculovírus ocorrem especialmente entre regiões flanqueadas por *hrs* (ARENDS & JEHLE, 2002; HAYAKAWA *et al.*, 2000), enquanto eventos de inserção e deleção de sequências têm sido observadas principalmente em regiões adjacentes a estes elementos repetitivos, onde são encontrados muitos dos genes *bro* (BLISSARD & ROHRMANN, 1990; COCHRAN & FAULKNER, 1983; HAYAKAWA *et al.*, 2000; KUZIO *et al.*, 1999).

Amostras selvagens de baculovírus apresentam grande diversidade genética intrapopulacional, onde diferentes genótipos virais coexistem e coinfectam populações de hospedeiros (BROWN *et al.*, 1985; COOPER *et al.*, 2003; GETTIG & MCCARTHY, 1982; HODGSON *et al.*, 2001; KISLEV & EDELMAN, 1982; LEE & MILLER, 1978; MAEDA *et al.*, 1990), as quais por sua vez também possuem uma diversidade genética intrínseca (PASHLEY & JOHNSON, 1986; SOSA-GÓMEZ, 2004). Em ensaios de infectividade *in vivo*, a patogenicidade de diferentes variantes virais frente a populações heterogêneas de hospedeiros (estimada como dose letal para 50% dos insetos (LD₅₀),) pode variar em até sete vezes entre os vários genótipos (HODGSON et al., 2001). Por outro lado, uma maior patogenicidade foi observada em infecções com genótipos misturados, se comparadas a infecções por um único genótipo. Em situações de coinfecção, os diferentes tecidos e células podem favorecer uns genótipos em detrimento de outros, o que promoveria a compartimentalização dos recursos entre os diferentes genótipos, evitando a competição direta e levando a uma rápida colonização e consequente morte dos hospedeiros (HODGSON et al., 2004). A diversidade em populações selvagens de AgMNPV já foi documentada em estudos anteriores, onde diferentes *loci* variáveis foram identificados. No entanto, nem todos elementos genéticos impactados por estas variações foram caracterizados, e a influência destes polimorfismos sobre a evolução viral ainda não foram esclarecidas (CROIZIER & RIBEIRO, 1992; MARUNIAK, 1989; MARUNIAK *et al.*, 1999).

1.5 BACULOVÍRUS Anticarsia gemmatalis MNPV

O primeiro registro de um NPV de *Anticarsia gemmatalis* ocorreu no Peru, em 1962, onde lagartas com sintomas de poliedrose nuclear foram detectadas em cultivos de alfafa (STEINHAUS & MARSH, 1962). Em 1972, em Campinas-SP, ocorreu o primeiro registro de um AgMNPV em território brasileiro. Análises de infectividade *in vivo* a partir desta amostra apontaram a alta patogenicidade deste nucleopoliedrovírus contra a lagarta-da-soja (*Anticarsia gemmatalis* Hubner), o que colocou este vírus como um eficiente agente de controle biológico de populações deste lepidóptero (ALLEN & KNELL, 1977). Amostras desta população viral detectada em Campinas foram utilizadas para controle biológico na Florida (EUA) em 1979. A partir de larvas mortas por variantes selvagens derivados dos vírus aplicados neste ano (AgMNPV-79), seis variantes foram isolados por *plaque assay*, o que resultou na obtenção de um protótipo deste nucleopoliedrovírus, o isolado AgMNPV-2D, o qual representava o genótipo predominante (1/3) desta população viral (JOHNSON & MARUNIAK, 1989; MARUNIAK, 1989).

No final da década de 70, variantes de AgMNPV derivados diretamente do isolado de Campinas (1972) estavam sendo utilizados em cultivos de soja na Florida, enquanto no Brasil outro isolado, obtido em Londrina, estava sendo empregado em um programa piloto de controle biológico conduzido pela CNPSo-EMBRAPA em meados de 1980 (MOSCARDI, 1989). A multiplicação viral de AgMNPV em lagartas *A. gemmatalis* por anos seguidos em condições ambientais distintas (tanto climáticas como de hospedeiros) proporcionaram a fixação de polimorfismos com origens independentes nestas duas populações de AgMNPV (MARUNIAK, 1989).

No Brasil, entre os anos de 1982 e 1985, a extensão da área de cultivos de soja tratados com baculovírus cresceu 100 vezes, isto é, de 2.000 para 200.000 ha. Com a implementação de unidades de produção de baculovírus no sul do país, um isolado obtido em 1985 (AgMNPV-85) passou a ser produzido em larga escala, fator que fez com que a área plantada tratada com baculovírus subisse para 500.000 ha em 1988 (MARUNIAK et al., 1999; MOSCARDI, 1989). Em termos de extensão, o auge do programa de controle biológico com o baculovírus *Anticarsia* se deu na safra 2003/2004, onde cerca de 2 milhões de hectares foram tratados com o vírus, representando 10% da área plantada com soja naquele ano no Brasil (MOSCARDI et al., 2011). Testes comparativos entre áreas tratadas com baculovírus; com inseticidas químicos; e não tratadas, revelaram que AgMNPV foi capaz de reduzir em 80% a população de *A. gemmatalis*, alcançando o mesmo rendimento demonstrado pelo inseticida químico (MOSCARDI, 1999; MOSCARDI et al., 2011). O sucesso na utilização de AgMNPV no controle da lagarta-da-soja no Brasil fez dele o bioinseticida viral mais amplamente utilizado no mundo (MOSCARDI, 1999; OLIVEIRA et al., 2006).

Diferentes variantes virais utilizados em controle biológico tiveram seus genomas comparados por perfis de restrição do DNA. Variantes purificados a partir dos isolados 79 (Florida) e 85 (Londrina) evidenciaram a ocorrência de diversas inserções e deleções no genoma destes vírus ao longo de suas histórias evolutivas independentes (MARUNIAK, 1989; MARUNIAK et al., 1999). Os isolados geográficos analisados neste trabalho de mestrado correspondem a populações virais de ocorrência natural em cultivos de soja sem histórico de uso de baculovírus como agente de controle biológico (MOSCARDI, comunicação pessoal)¹. Neste sentido, muitas das variações genômicas identificadas nos isolados geográficos podem corresponder a alterações ocorridas na linhagem viral que deu origem ao isolado AgMNPV-85 e seus descendentes. Os polimorfismos presentes nos variantes do isolado 85 não foram caracterizadas em detalhes, e tais variações certamente não estão presentes nos genomas dos variantes de AgMNPV-79, isolado de origem do protótipo AgMNPV-2D (MARUNIAK, 1989; MARUNIAK et al., 1999). O genoma do protótipo 2D possui 132.239, bp e codifica 152 ORFs, das quais três são exclusivas deste variante (Figura 5) (OLIVEIRA et al., 2006).



Figura 5 – Mapa circular do genoma do AgMNPV-2D.

As setas vermelhas e azuis indicam ORFs em senso positivo e negativo, respectivamente. As regiões em amarelo representam regiões homólogas (hr). O gráfico circular mais interno ilustra o conteúdo GC (%) ao longo do genoma.

Fonte: Modificado de (OLIVEIRA, 2010), com permissão.

Análises filogenéticas entre AgMNPV e outros baculovirus classifica este vírus entre os alfabaculovírus do grupo I (ZANOTTO *et al.*, 1993), com grande proximidade filogenética com CfDefMNPV, com o qual possui similaridade de sequências de proteínas superior a 92% (OLIVEIRA et al., 2006) (Figura 6). Uma das principais diferenças entre esses dois vírus é a ausência dos genes *chitinase* e *cathepsina* em AgMNPV, genes cujos produtos podem ser responsáveis pela liquefação do inseto infectado. Esta característica fenotípica não é observada em infecções por AgMNPV, o que favorece a coleta dos insetos mortos para reaplicação futura, fator que explica parte do sucesso deste vírus como agente de controle biológico (OLIVEIRA et al., 2006).

Figura 6 – Relações filogenéticas entre os baculovírus dos quatro gêneros: *Alphabaculovirus*; *Betabaculovirus*; *Gammabaculovirus*; e *Deltabaculovirus*.



Em destaque (região laranja), o AgMNPV, aqui representado pelo isolado 2D, inserido no grupo I entre os alfabaculovírus. Fonte: Modificado de (OLIVEIRA, 2010), com permissão.

O presente trabalho visou o sequenciamento de 17 populações selvagens de AgMNPV. As análises que serão realizadas possibilitam a identificação de polimorfismos em regiões codificantes e regulatórias. As diferentes observadas entre os diferentes haplótipos presentes nestas populações heterogêneas trarão importantes evidências sobre o modo de evolução deste vírus em ambiente natural, onde esta espécie viral foi progressivamente utilizada ao longo das décadas de 80, 90 e 2000 como agente de controle biológico da lagarta-da-soja. As informações levantadas neste trabalho serão de importância ímpar para a retomada do uso deste vírus como bioinseticida.

2 OBJETIVOS

Este trabalho tem como objetivo geral seqüenciar e analisar os genomas de 17 populações selvagens do nucleopoliedrovírus *Anticarsia gemmatalis* (AgMNPV) coletadas em diferentes regiões geográficas da América do Sul, no período de 1980 a 1990. Entre os objetivos específicos deste trabalho, estão:

- (i) Identificar polimorfismos em geral e em seqüências codificantes e reguladoras da expressão gênica de AgMNPV em particular, tendo como referência o genoma do protótipo AgMNPV-2D;
- (ii) Avaliar as diferenças de organização e composição gênica na metapopulação de AgMNPV utilizando métodos comparativos;
- (iii) Esclarecer mecanismos de variação genômica em isolados selvagens;
- (iv) Identificar outras possíveis autapomorfias e distinções genômicas específicas, além da *parp-like* (ORF 31) do AgMNPV-2D;
- (v) Obter estimativas de diversidade genética interpopulacional que possam indicar particularidades relevantes a cerca da variabilidade genética em AgMNPV.

3 MATERIAIS E MÉTODOS

3.1 AMOSTRAS VIRAIS

Este estudo envolveu 17 amostras de populações virais selvagens de AgMNPV isoladas a partir de lagartas de *Anticarsia gemmatalis* infectadas pelo vírus. Estas são populações virais de ocorrência natural em cultivos de soja sem histórico anterior de aplicação de baculovírus como agente de controle biológico. As amostras foram coletadas entre os anos de 1980 e 1990 (Tabela 1), nos estados brasileiros de Mato Grosso do Sul, Paraná e Rio Grande do Sul, além de Uruguai e Argentina (Figura 7). Coordenadas geográficas precisas dos locais de origem não estão disponíveis. As amostras virais foram gentilmente cedidas pelo Dr. Daniel Sosa-Gomez (Embrapa Soja - Londrina, PR).

Isolado	Localização	Ano
NPV26	MS - Dourados	1984
NPV27	AR - Argentina	1982
NPV28	UY - Uruguai	1984
NPV29	PR - Ponta Grossa	1985
NPV30	UY - Uruguai	1982
NPV31	PR - Sertanópolis	1986
NPV32	MS - Dourados	1982
NPV33	RS - Pelotas	1984
NPV34	PR - Londrina (Distrito da Warta)	1990
NPV35	RS - Rio Grande do Sul	1984
NPV36	RS - Passo Fundo	1982
NPV37	PR - Campo Mourão	1982
NPV38	UY - Uruguai	1982
NPV39	PR - Sertanópolis	1987
NPV40	PR - Londrina	N/A^1
NPV42	N/A ² (Isolado de laboratório)	1987
NPV43	PR - Londrina (Prop. Luis Pesce)	1980

Tabela 1 – Lista de isolados geográficos de AgMNPV, localização e ano de coleta.

¹ ano de coleta indeterminado;

² região geográfica de origem não especificada.



Figura 7 - Regiões geográficas de origem dos isolados de AgMNPV (localizações aproximadas).

3.2 AMPLIFICAÇÃO DOS ESTOQUES VIRAIS IN VIVO

As amostras virais foram amplificadas por meio de infecções *per os* em lagartas de *Anticarsia gemmatalis* de 3° ou 4° ínstares (O'REILLY *et al.*, 1993). Os insetos foram mantidos em 18 potes plásticos de 500 mL, com tampa e entrada de ar protegida com algodão (Figura 8A; Figura 8B). Cada pote recebeu 50 insetos e 10 blocos de dieta artificial (0,5 cm³) (GREENE *et al.*, 1976) inoculados com 20 µl de suspensão de corpos de oclusão poliédricos (OB) (250 OBs/µl): uma média de 1000 OBs/lagarta. Após a completa ingestão dos blocos inoculados, no segundo dia de experimento, os insetos foram providos com dieta livre de vírus (5 blocos de 1,5 cm³/pote). Um pote foi mantido em mesmas condições, porém sem a aplicação de inóculo infeccioso, para efeito de controle negativo (Figura 8C)

A amostra NPV29 continha baculovírus em sua forma extracelular (*Budded virus*). Para obter estoques virais deste isolado foi procedida uma infecção intrahemocélica, com uso de uma seringa de 1 ml (agulha para aplicação de insulina) (JONES, 2000). O estoque inicial foi diluído em proporção 1:10, e cada lagarta entre 4° e 5° ínstar foi injetada com 25 μ l de inóculo via terceiro par de pseudopatas (Figura 9). Anteriormente à injeção, os insetos foram anestesiados em gelo por 15 min.. Ao todo, cinco lagartas foram infectadas e colocadas em um pote de 500 ml com 10 cubos de dieta (1,5 cm³).

Entre o 7° e o 10° dia pós-infecção (Figura 8D), as lagartas mortas com sintomas de poliedrose nuclear (em ambos os experimentos) foram coletadas e mantidas a -20 °C.



Figura 8 – Amplificação das populações virais em larvas de A. gemmatalis.

A) Potes de criação de lagartas B) Detalhe de um pote de criação no início do processo de amplificação, com lagartas de 3º e 4º ínstares. C) Pote controle contendo lagartas não infectadas. Ao centro observa-se a presença de pupas e lagartas em estágio pré-pupa. D) Pote contendo lagartas infectadas. Ao centro observa-se uma lagarta com o aspecto típico de poliedrose nuclear (clareamento da epiderme).



Figura 9 - Injeção intrahemocélica de inóculo viral de Budded Virus em larva de A. gemmatalis.

Cerca de 5 mm da agulha foi inserida na longitudinal entre o terceiro par de pseudopatas (seta), buscando não perfurar o intestino do inseto, a fim de inocular as partículas virais exclusivamente na hemocele.

3.3 PURIFICAÇÃO DOS CORPOS DE OCLUSÃO

A partir das larvas infectadas, foi procedida a purificação dos vírus para a extração do DNA genômico viral. As lagartas foram maceradas com tampão de homogeneização (2 ml/g de inseto) e as amostras foram filtradas em duas camadas de gaze de algodão para remoção dos restos de tecido dos hospedeiros. O macerado foi centrifugado a 9000 RPM por 15 min. a 4 °C (em Sorvall RC-5C Plus, rotor SA-600); os sobrenadantes foram descartados; os "*pellets*" coletados e ressuspendidos em 20 ml de tampão TE (pH 7,6). As amostras foram submetidas a ultra-som (sonicador) por dois ciclos de 30 segundos intercalados por 30 segundos e mantidas em gelo. Em seguida foram novamente centrifugadas, desta vez a 11000 RPM por 12 min. a 4 °C (em Sorvall RC-5C Plus, rotor SA-600). Os sobrenadantes foram descartados e os "*pellets*" novamente ressuspendidos em 20 ml de tampão TE.

Cerca de 5 ml do material obtido ao fim das centrifugações foram aplicados sobre gradientes de sacarose (40 a 65%) previamente preparados em tubos de 35 ml. Os tubos foram ultracentrifugados a 24000 RPM por 3 horas a 25°C (Beckman, rotor SW28). Os sobrenadantes foram descartados, os "*pellets*" foram diluídos (1:4) com tampão TE e então centrifugados a 11000 RPM por 15 min. a 4°C (em Sorvall RC-5C Plus, rotor SA-600). O material final foi ressuspendido em 50 ml de água MilliQ e armazenado a -20 °C (MARUNIAK, 1986).

3.4 EXTRAÇÃO DE DNA DAS AMOSTRAS VIRAIS

Inicialmente, 1 ml de suspensão contendo OBs purificados (concentração indefinida) foi dissolvida em 500 ml de solução alcalina DAS: 100 mM NaHCO₃.Na₂CO₃, 50 mM NaCl, 2 mM Na-EDTA (pH 10,5) (WHITT & MANNING, 1987). O material foi mantido a 37 °C por 30 minutos em termobloco. Em seguida o volume foi centrifugado por 15 minutos a 1000 g e o sobrenadante transferido para um novo tubo com 100 µl de tampão Tris.Cl (pH 6,8). Em seguida as amostras foram centrifugadas a 30000 g por 10 minutos, o sobrenadante foi descartado e 200 µl de tampão de disrupção (10 mM Tris (pH 7.6); 10 mM EDTA; 0,25% SDS) foram adicionados juntamente com 15 µl de proteinase K (25 mg/ml). A mistura de reação foi incubada por 4 horas a 56 °C. Em seguida o DNA foi extraído seguindo um protocolo padrão de Fenol:Clorofórmio:Álcool isoamílico (25:24:1) e precipitado com etanol 70% (*overnight*). Após secagem do etanol em centrífuga a vácuo (*Eppendorf Vacufuge Concentrator*[®], *Eppendorf*), o "pellet" foi eluído em 50 µl de tampão TE (pH 7,6) e quantificado por espectrofotometria (*Spectrophotometer ND-1000*[®], NanoDrop).

Para maximizar a cobertura do sequenciamento de material viral as alíquotas de suspensão de vírus encapsidados foram tratadas com DNAse para eliminar eventuais contaminações por DNA celular.

3.5 PIROSEQUENCIAMENTO

As amostras de DNA genômico dos diferentes isolados foram preparadas para sequenciamento. Cerca de 1,5 µg de DNA viral de cada isolado foram enviadas para pirosequenciamento em uma plataforma Roche 454[®] da *University Medical Center Göttingen* (Alemanha), em colaboração com o laboratório Prof. Dr. Manfred Weidmann e do Dr. Meik Dilcher.

Inicialmente, 50 ng de DNA de cada amostra foram utilizados em reações de preparação de bibliotecas. Para isso, foi utilizado o *Nextera™ DNA Sample Prep Kit* (Epicentre *Biotechnologies*), o qual utiliza mecanismos de transposição para fragmentar e marcar as moléculas de DNA. A reação de transposição *in vitro* ocorre pela ação de enzimas transposases em associação com extremidades de elementos transponíveis. Após incubação, os *transposons* são incorporados em diversas regiões do DNA-alvo, proporcionando a fragmentação e ligação covalente dos oligonucleotídeos *transposons* nas extremidades 5'.

Além da transposição e fragmentação, as moléculas de DNA de cada uma das 17 amostras também foram marcadas com diferentes MIDs — *Multiplex Identifiers, GS FLX Titanium Rapid Library MID Adaptors Kit*TM (454 *Life Sciences*, Roche) — e ligados a adaptadores. Os adaptadores servem de suporte para imobilização e amplificação dos fragmentos na etapa de PCR em emulsão (emPCR); e as MIDs para posterior identificação das amostras *in silico* (Tabela 2).

Subconjuntos	Amostras	Marcadores
	NPV26	MID1
	NPV27	MID2
	NPV28	MID3
	NPV29	MID4
	NPV30	MID5
	NPV31	MID6
1° subconjunto	NPV32	MID7
	NPV33	MID8
	NPV34	MID9
	NPV35	MID10
	NPV36	MID11
	NPV37	MID12
	NPV38	MID1
	NPV39	MID2
2º subconjunto	NPV40	MID3
2 subconjunto	NPV42	MID4
	NPV43	MID5

Tabela 2 – Lista de marcadores (barcodes) utilizados para identificação das amostras in silico.

Em seguida as bibliotecas foram purificadas com *Agencourt AMPure XP*TM (Beckman Coulter, Inc) para a eliminação de fragmentos menores que 300 bp. Nesta etapa, 50 μ l de tampão TE foram adicionados a 50 μ l de biblioteca *Nextera*TM e 70 μ l de *AMPure XP* (*beads* paramagnéticas), resultando numa proporção 1:0,7 (DNA:*beads*). Esta mistura foi incubada em um *rotator* por 10 minutos a temperatura ambiente. Os tubos foram submetidos a um *DynaMag*TM-2 *magnet* (Life Technologies Corporation) para isolamento das *beads* paramagnéticas associadas a DNA. O material capturado foi lavado 3 vezes com 500 μ l de etanol 70%, secadas em temperatura ambiente por 3 minutos e ao final foi eluído em 12 μ l de tampão TE. O procedimento de construção de bibliotecas gera conjuntos de fragmentos que medem entre 500 e 1100 bp, e a distribuição de tamanhos de fragmentos pode ser controlada
mudando-se a quantidade de *transposons* e o tempo de reação na etapa inicial. Um μ l de cada amostra foi analisado em um *High Sensitivity DNA Chip*TM de um 2100 *Bioanalyzer*TM (Agilent *Techonologies*) para avaliar a distribuição de tamanhos dos fragmentos e quantificar as bibliotecas, as quais posteriormente foram diluídas a 1 × 10⁷ moléculas/ μ l.

Foram preparados dois subconjuntos de amostras contendo 20 µl de cada amostra. O primeiro subconjunto foi constituído por 14 amostras, e o segundo pelas outras cinco amostras, juntamente com 3 amostras de *Bacillus spp* (Tabela 2). Estes dois subconjuntos foram diluídos a uma proporção 1:10 em tampão TE. Seguindo orientações do fabricante (*emPCR Method Manual – Lib-L SV*, Roche) este material foi utilizado para titulação, objetivando-se obter a proporção correta entre moléculas de DNA (dos subconjuntos 1 e 2) e *beads* de pirosequenciamento. A titulação mostrou que a proporção ideal para cada subconjunto foi 2 cópias por *bead* (subcojunto 1) e 16 cópia por *bead* (subconjunto 2). Em seguida o material foi submetido a uma emPCR para grande volumes, seguindo instruções do fabricante (*emPCR Method Manual – Lib-L LV*, Roche). A emPCR mostrou um rendimento de 13,7% para o subconjunto 1, e 6,4% para o subconjunto 2. Uma *PicoTiterPlate* foi dividida em 2 regiões, e cada uma delas individualmente recebeu 2×10^6 *beads* associadas a DNA dos dois subconjuntos. O material foi pirosequenciado em uma plataforma *GS FLX Titanium*[®] (454 *Life Sciences*, Roche) seguindo recomendações do fabricante (*Sequencing Method Manual, GS FLX Titanium Series*, Roche).

3.6 REMOÇÃO DOS ADAPTADORES

Os arquivos de saída do sequenciamento foram disponibilizados no formato SFF (*Standard Flowgram Format*), em arquivos individuais contendo as leituras (*reads*) dos diferentes isolados. Após sequenciados, cada leitura é composta pela sequência de interesse na porção central, a qual é flanqueada pelos adaptadores e pelas MIDs. Após a segregação das sequências de acordo com suas MIDs, estas são removidas, e o usuário final recebe leituras contendo apenas as sequências de interesse ligadas aos adaptadores nas extremidades (Figura 10).





Inicialmente os arquivos foram convertidos ao formato FASTQ por meio do programa *sff_extract* 0.2.13 (BLANCA & CAÑIZARES, 2012). Durante esta etapa de conversão também foi realizada a remoção das sequências dos adaptadores (34 bp) da extremidade 5'. A remoção dos adaptadores da extremidade 3' foi executada posteriormente, utilizando-se o programa *cutadapt*, em três iterações (MARTIN, 2011).

3.7 MAPEAMENTO DAS LEITURAS SOBRE O GENOMA REFERÊNCIA

Anteriormente a montagem dos genomas, cada conjunto de leituras dos diferentes isolados geográficos foi caracterizado quanto a localização e identidade de cada sequência em relação ao genoma de referência do vírus AgMNPV-2D (número de acesso GenBank/EMBL/DDBJ: DQ813662). Um *script* Perl, denominado *topHSPreport.pl*, foi desenvolvido para implementar buscas por meio do programa BLASTN (nos parâmetros padrão) e gerar um relatório final contendo apenas um único resultado de HSP (*high-scoring segment pairs*) para cada leitura. Desta maneira, caso alguma leitura tenha correspondência com mais de uma região, e consequentemente tenha mais de um HSP, como em situações de regiões duplicadas no genoma, apenas as informações do HSP mais relevante, isto é, o de maior *Score* e menor e-*value*, são mantidas para as análises subsequentes (Tabela 3).

Tabela 3 – Dados relativos aos HSPs, em formato de saída tabular, obtidos para 10 leituras do isolado NPV35. O algoritmo implementado no script *topHSPreport.pl* seleciona e reporta apenas os dados do único HSP mais relevante de cada leitura (destacados em *itálico*). Desta forma, os resultados menos significativos (em negrito) são desconsiderados.

#	Leitura	Subject	% de identidade	Tamanho do HSP	Mis- match	Gap	<i>Query</i> início	Query fim	Subject início	<i>Subject</i> fim	e-value	Score
1	HQSC8GL01A0MOM	AgMNPV	99.53	424	2	0	1	424	68572	68995	0.0	773
2	HQSC8GL01A0MSJ	AgMNPV	100.00	426	0	0	1	426	39820	40245	0.0	787
3	HQSC8GL01A0MTY	AgMNPV	93.31	239	7	5	231	462	21500	21264	1,00E-96	344
4	HQSC8GL01A0MXN	AgMNPV	85.68	412	52	7	13	422	7028	7434	1,00E-121	427
4	HQSC8GL01A0MXN	AgMNPV	82.74	226	32	7	200	422	26497	26276	1,00E-51	195
5	HQSC8GL01A0N3K	AgMNPV	<i>98.43</i>	381	4	2	1	379	72079	72459	0.0	669
6	HQSC8GL01A0N45	AgMNPV	100.00	407	0	0	1	407	117259	117665	0.0	752
7	HQSC8GL01A0N4K	AgMNPV	98.93	375	0	4	1	375	84937	85307	0.0	667
8	HQSC8GL01A0N6J	AgMNPV	99.72	351	0	1	1	350	3184	3534	0.0	641
8	HQSC8GL01A0N6J	AgMNPV	90.84	131	12	0	215	345	2656	2786	3,00E-46	176
9	HQSC8GL01A0N8F	AgMNPV	81.78	258	28	13	96	350	91578	91337	8,00E-53	198
9	HQSC8GL01A0N8F	AgMNPV	91.30	115	8	2	293	406	25590	25477	5,00E-40	156
9	HQSC8GL01A0N8F	AgMNPV	85.47	117	12	4	36	149	91441	91327	2,00E-28	117
9	HQSC8GL01A0N8F	AgMNPV	80.62	160	14	14	1	157	91342	91197	1,00E-25	108
10	HQSC8GL01A0NA6	AgMNPV	97.89	332	2	4	1	327	81229	81560	1,00E-164	569

Em seguida, os dados dos HSPs mais relevantes foram plotados em planilhas do programa *Microsoft Office Excel 2007* (em português). Informações adicionais a cerca das leituras foram obtidas (Tabela 4) com o emprego de "funções de pesquisa e referência" (PROC, PROCV); "funções lógicas" (E, SE); "funções estatísticas" (MÁXIMO, MÍNIMO, MÉDIA); "funções matemáticas e trigonométricas" (ARREDONDAR.PARA.CIMA, INT); além de operações matemáticas básicas. Os valores associados ao tamanho das leituras gerados nos resultados do BLASTN referem-se ao tamanho de cada HSP, e não ao tamanho das leituras em si. Para se obter o tamanho real de cada leitura, um *script* Perl, denominado *numbases.pl*, foi desenvolvido para reportar o número de bases de cada uma das leituras dos 17 conjuntos de dados.

		111 1 35	upresente	idub IIu I	uoonu 5.				
#	Cobertura da leitura	Bases não cobertas	N ^{o1} da u <i>Feature</i> ²	Nº da d <i>Feature</i> ³	Nome da u <i>Feature</i>	Nome da d <i>Feature</i>	Ìndice ⁴	Extensão ⁵	Cabeçalho modificado
1	100,00%	0	82	83,5	vp91	hr2b	082	3	082_3_HQSC8GL01A0MOM
2	100,00%	0	49	50	odve66	etm	049	2	049_2_HQSC8GL01A0MSJ
3	50,22%	230	25	25	39K	39K	unc	1	unc_1_HQSC8GL01A0MTY
4	97,16%	12	11	11	bro-c	bro-c	bro	1	bro-c_1_HQSC8GL01A0MXN
5	100,00%	0	87	87	lef4	lef4	087	1	087_1_HQSC8GL01A0N3K
6	100,00%	0	136	137	me53	ie0	136	2	136_2_HQSC8GL01A0N45
7	100,00%	0	100	100	<i>p</i> 87	<i>p</i> 87	100	1	100_1_HQSC8GL01A0N4K
8	100,00%	0	6	7	bro-a	bro-b	bro-a	2	bro-a_2_HQSC8GL01A0N6J
9	62,81%	151	104,5	104,5	hr3	hr3	unc	1	unc_1_HQSC8GL01A0N8F
10	100,00%	0	95	96	lef5	p6.9	095	2	095_2_HQSC8GL01A0NA6

 Tabela 4 – Exemplo de informações adicionais obtidas para as mesmas 10 leituras do isolado

 NPV35 apresentadas na Tabela 3.

¹Número da ORF segundo descrito em (OLIVEIRA et al., 2006);

² Upstream Feature: ORF ou hrs onde a leitura inicia sua correspondência;

³ Downstream Feature: região genômica onde a leitura termina sua correspondência;

⁴Código identificador da leitura;

⁵ Número de *Features* que a leitura cobre.

Como mostrado na Tabela 4, cada leitura foi mapeada no genoma de referência de forma a se obter informações sobre cobertura, assim como o nome e/ou número das ORFs ou *hrs* correspondentes ao início (u*Feature*; u = *upstream*) e o fim (d*Feature*; d = *downstream*) de cada leitura. Regiões não-codificantes, como *hrs* e *drs*, foram numeradas com o número inteiro referente à ORF (*Open Reading Frame*) que a precede, seguido de um valor decimal (Ex.: hr2b = 83,5).

O objetivo final do mapeamento foi gerar um novo cabeçalho para cada sequência, elemento essencial dentro da abordagem de montagem de genomas adotada neste trabalho (vide seção 3.8). Tendo como referência a orientação em sentido 5' \rightarrow 3' da fita senso positivo do vírus AgMNPV-2D, cada leitura que correspondesse a alguma região do genoma referência recebeu um índice (código identificador) em formato de três digitos (Ex. 082); um número em referência extensão da leitura, isto é, a quantidade de *features* que a leitura cobre; seguido do nome original da leitura (vide última coluna da Tabela 4).

O código identificador de cada leitura foi definido da seguinte forma:

i. Leituras que cobriam uma única *feature* foram identificadas com o número que esta *feature* possui segundo descrito em (OLIVEIRA et al., 2006) (Figura 11A).

Caso a leitura correspondesse há uma *feature* que possui uma sequência que se repete ao longo do genoma, como genes *bro* e *hrs*, esta era identificada pelos códigos "bro" e "hr", respectivamente;

- *ii.* Leituras que cobriam duas *features* foram identificadas com o número da primeira *feature* a montante (u*Feature*) (Figura 11B);
- *iii.* Leituras que cobriam três *features* foram identificadas com o número da *feature* central (Figura 11C);
- *iv.* Leituras que cobriam quatro *features* foram identificadas com o número da segunda *feature* a montante (Figura 11D).
- v. Leituras que iniciavam em "regiões intergênicas" foram identificadas com o número da *feature* a montante no genoma (Figura 11E);
- vi. Leituras com cobertura parcial isto é, que apresentavam nas extremidades mais de 25 bases consecutivas divergentes em relação à referência – foram identificadas com o código "unc" (*uncovered*) (Figura 11F).
- *vii.* Leituras que não correspondiam a nenhuma região do genoma referência foram mantidas com seus nomes originais.

Figura 11 - Identificação de leituras de pirosequenciamento em relação a um genoma referência hipotético.



A) Leitura que coincide internamente a uma única ORF (41). B) Leitura que inicia sua coincidência na ORF41 e se estende até a ORF 42. C) Leitura que iniciasse na ORF41, perpassa a ORF 42, e se estende até o início da ORF43. D) Leitura que se inicia no fim da ORF41, perpassa duas ORFs antes de terminar no início da ORF44. E) Leitura que se inicia numa região intergênica e termina sua coincidência na próxima ORF a jusante. F) Leitura que possui uma região para a qual o BLASTN não reportou homologia com a ORF 43, e que termina sua coincidência na ORF44.

O mapeamento das leituras frente ao genoma de AgMNPV-2D foi realizado individualmente para cada um dos 17 conjuntos de dados de pirosequenciamento. A renomeação das leituras em formato FASTQ foi realizada por meio do *script* Perl denominado *seqrenameQ.pl*.

3.8 RECONSTRUÇÃO DOS GENOMAS PREDOMINANTES

3.8.1 Montagens preliminares

Anteriormente a reconstrução efetiva dos genomas, foram realizadas montagens preliminares utilizando o programa *MIRA assembler* 3.4.0 (CHEVREUX *et al.*, 1999). Essas montagens envolveram todas as leituras obtidas para cada isolado, e se deu adotando-se duas abordagens diferentes. A primeira foi configurada para realizar montagens menos restritas, com o intuito de reconstruir sequências consenso. O objetivo dessa abordagem foi avaliar a ocorrência de possíveis rearranjos estruturais presentes nos genomas de cada população. A segunda abordagem visou à reconstrução dos diferentes haplótipos presentes nas populações, em especial aqueles relativos a regiões repetitivas (isto é, duplicadas, como os genes *bro*), e regiões repetidas em tandem ao longo do genoma (como as *hrs*). Para isso, utilizou-se uma parametrização mais rigorosa, que permitiu apenas o agrupamento de leituras 100% idênticas, garantindo a reconstrução isolada dos diferentes haplótipos (Tabela 5). Os *contigs* montados nesta segunda abordagem foram utilizados para preencher possíveis lacunas entre *contigs* que colapsaram em regiões com repetições em tandem, como regiões homólogas (*hrs*) e *direct repeats (dr*), as quais são de difícil reconstrução.

Comando MIRA	Função do parâmetro	Parametrização menos restrita	Parametrização mais restrita
-highlyrepetitive	Reconstrói regiões repetitivas por meio de múltiplas iterações	yes	yes
-pr	Controla a quantidade de coincidências exatas (em pares de base) que deve ser inicialmente alcançada para considerar uma sobreposição como uma coincidência válida para gerar um alinhamento entre leituras	25 bp	25 bp
-urd	Distribui as leituras uniformemente durante a montagem (procedimento adequado para a reconstrução de regiões repetidas em tandem)	yes	yes
-sd	Detecta leituras que por ventura ainda possuam adaptadores e as remove caso estejam causando perturbações ao alinhamento que impeça a junção de <i>contigs</i> pequenos em <i>contigs</i> maiores	yes	yes
-mrl	Utiliza apenas leituras com o tamanho mínimo especificado	25 bp	25 bp
-mrpc	Define o número mínimo de leituras que um contig deve ter para ser construído e salvo	по	2
-mrs	Alinha apenas leituras que apresentem o mínimo de identidade especificada	55%	100%
-ace	Gera alinhamentos no formato ACE	yes	yes

Tabela 5 – Parâmetros e valores utilizados para as montagens com o programa MIRA.

Para avaliar graficamente a qualidade dos alinhamentos, o programa *Consed* 23.0 (GORDON *et al.*, 1998) foi utilizado para a visualização dos arquivos em formato ACE gerados após as montagens.

3.8.2 Setorização do genoma de AgMNPV-2D

Com o objetivo de facilitar o processo de montagem, o genoma referência AgMNPV-2D foi estrategicamente dividido em quatro setores genômicos medindo entre 30040 e 36367 bp (Figura 12).



Figura 12 – Setorização do genoma de AgMNPV-2D.

Na região de intersecção entre cada um dos setores estão localizadas ORFs com mais de 750 bp; baixa diversidade genética (isto é, com alto grau de conservação); e que estão relativamente eqüidistantes umas das outras no genoma, são elas: ORF19 (*lef1*); ORF62 (*lef9*); ORF99 (*p48*) e ORF138 (*p49*). A fim de reconstruir o genoma completo, estas ORFs serviram de ponto de ancoragem para a junção dos *contigs* gerados para os diferentes setores. Partindo-se da orientação em sentido 5' \rightarrow 3' da fita senso do vírus AgMNPV-2D, cada setor inicia-se na primeira base a montante e se estende até a última base a jusante das ORFs conservadas que os delimitam.

3.8.3 Montagem dos genomas

De posse das leituras devidamente identificadas, as montagens dos genomas foram executadas por meio do programa *Aligner (CodonCode Corp.)*. Para reconstruir cada um dos setores de maneira isolada, apenas o subconjunto de leituras que contém sequências correspondentes às ORFs presentes em cada setor eram incluídas em cada ciclo de montagem:

- *i.* Setor A: leituras com índice maior ou igual a 137 e menor ou igual a 019;
- *ii.* Setor B: leituras com índice maior ou igual a 018 e menor ou igual a 062;
- *iii.* Setor C: leituras com índice maior ou igual a 061 e menor ou igual a 099;
- *iv.* Setor D: leituras com índice maior ou igual a 098 e menor ou igual a 138.

Adicionalmente, além dos subconjuntos específicos de cada setor, todas as leituras que possuíam os índices "bro" ou "hr", assim como leituras sem homologia detectada com AgMNPV-2D via BLASTN, foram reiteradamente incluídas em cada ciclo de montagem dos diferentes setores. A inclusão sistemática destas leituras nas montagens visava garantir a reconstrução de regiões repetitivas ao longo do genoma (genes *bro* e *hrs*), assim como possíveis regiões não presentes no isolado 2D, portanto exclusivas de cada isolado. Leituras com os índices 137, 018, 061 e 098 foram adicionadas as montagens dos setores A, B, C e D, respectivamente, para garantir a inclusão das sequências das ORFs 138, 19, 62 e 98 que por ventura estejam em leituras que se iniciam em uma ou duas ORFs a montante (Figura 11). Para a montagem, foram utilizados parâmetros de alinhamento específicos (Tabela 6).

Parâmetros	Valores
Algoritmo	End to end alignments
Porcentagem mínima de identidade	97%
Comprimento mínimo de sobreposição	150
Pontuação mínima	20
Porcentagem máxima de extremidades não alinhadas	5%
Tamanho máximo de lacunas	6
Tamanho da palavra (Word length)	24
Número máximo de falhas sucessivas	50
Pontuação de coincidência	1
Penalidade por mau pareamento	-2
Penalidade por lacuna	-2
Penalidade por lacunas adicionais	-3

Tabela 6 – Valores dos parâmetros empregados durante a montagem inicial dos setores genômicos.

Ao fim das montagens, apenas os *contigs* com 50 ou mais leituras foram considerados para a reconstrução de cada setor. Os *contigs* com maior número de sequências associadas reconstroem os haplótipos mais frequentes, os quais quando unidos numa única estrutura genômica, reconstroem o provável genótipo predominante de cada população (Figura 13).



Figura 13 – Modelo de reconstrução do possível genótipo predominante de cada população de AgMNPV.

numa estrutura genômica única recupera o provável genótipo mais frequente em cada população.

3.8.4 Junção de contigs dos setores genômicos

Devido à presença de regiões polimórficas, e em razão da existência de inserções em regiões homopoliméricas, parâmetros de montagem gradualmente menos restritos foram adotados para garantir a junção dos *contigs* de cada setor genômico. As regiões de junção entre *contigs* foram cuidadosamente inspecionadas com o intuito de identificar sobreposições demasiadamente divergentes ou não-homólogas, as quais poderiam resultar em estruturas genômicas artefactuais. Para contornar este problema, a localização genômica dos flancos dessas regiões duvidosas era identificada, e em seguida 3 ou mais novas leituras que sobrepunham ambos os flancos eram selecionadas a partir do conjunto de dados original, a fim de estabelecer a correta ligação genética entre os *contigs*. Quando a fragmentação dos *contigs* ocorria internamente a regiões codificantes, os SNPs suportados por mais leituras eram mantidos em detrimento dos SNPs menos freqüentes. A título de reconstrução do genótipo predominante na população, em situações de existência de múltiplas formas de um mesmo *locus*, apenas o variante mais freqüente era mantido na estrutura genômica final.

Na representação ilustrada acima, um genoma hipotético apresenta polimorfismos estruturais em 6 de seus 10 *locus* gênicos. Os retângulos de A a J, numerados de 1 a 5, caracterizam cada um dos haplótipos reconstruídos. A junção dos haplótipos de maior frequência (indicados pelo número 1)

3.9 CORREÇÃO DE INSERÇÕES E DELEÇÕES EM REGIÕES CODIFICANTES

Previamente as anotações genômicas, as sequências de regiões codificantes anotadas em AgMNPV-2D foram mapeadas contra os 17 genótipos reconstruídos. Esse procedimento visou à localização e correção de erros resultantes do sequenciamento de regiões homopoliméricas, os quais geram inserções ou deleções que fatalmente colapsam os quadros de leitura. As presenças ou ausências dos *indels* foram individualmente verificadas, confirmadas junto aos *contigs* originais, e manualmente corrigidas quando necessário.

3.10 ANOTAÇÕES GENÔMICAS

As ORFs presentes em cada genoma foram localizadas utilizando-se a ferramenta *ORF Finder* (http://www.ncbi.nlm.nih.gov/gorf/gorf.html). As coordenadas genômicas de todas as ORFs reportadas pelo programa foram plotadas em planilhas Excel, e a partir delas foram gerados arquivos em formato GFF (*General Feature Format*). Estes arquivos eram então visualizados e editados no programa *Artemis* (RUTHERFORD et al., 2000). A anotação de todas as ORFs homólogas a AgMNPV-2D foram transferidas para os gemonas dos isolados. As ORFs de possíveis novos genes encontrados em regiões intergênicas também foram assinaladas. A existência de regiões codificantes sobrepostas também foi considerada, e todas as ORFs com até 73,5% de sobreposição e tamanho mínino de 111 bp foram anotadas.

3.11 IDENTIFICAÇÃO DE NOVAS ORFS POR ESPECTROMETRIA DE MASSA

Com o intuito de identificar novas possíveis ORFs em AgMNPV, dados de espectrometria de massa do proteoma estrutural do isolado 2D (BRACONI; ZANOTTO, em fase de elaboração)² foram utilizados. Para isso, inicialmente, estoques virais foram obtidos a partir de infecções a 1 MOI em cultivos de células das linhagens SF9 e UFLAG. Ambos os fenótipos (BV e ODV) foram purificados em ultracentrifugações em gradiente de sacarose. As diferentes partículas virais foram separadamente aplicadas em eletroforeses SDS-PAGE (1D) e por eletroforeses bidimensionais (2D) em triplicata, e todas as bandas e *spots* foram extraídos dos géis e digeridos com Tripsina. Ao final, os peptídeos digeridos foram analisados por espectrometria de massa LC ESI-Q-TOF (bandas) e MALDI Q-TOF (*spots*).

Por meio da ferramenta *ORF Finder*, 428 supostas ORFs não anotadas no isolado AgMNPV-2D foram identificadas, e as sequências de aminoácidos codificadas por estas ORFs foram utilizadas para compor um banco de dados a partir do qual o programa *Mascot* (Matrix Science, London, UK; version 2.2.2) foi empregado para realizar o mapeamento dos peptídeos identificados por espectrometria de massa. Os parâmetros do *Mascot* foram estabelecidos com tolerância de fragmentação de massa dos íons entre 0,05 e 0,5 Da. Para a identificação dos peptídeos foi estabelecido valores de probabilidade acima de 99%, e com no mínimo 2 espectros únicos de identificação em relação ao algoritmo *Peptide Prophet* (KELLER et al., 2002).

3.12 ANÁLISE DE DOMÍNIOS PROTÉICOS

Buscas no *Conserved Domain Database* (MARCHLER-BAUER et al., 2011) foram realizadas para a identificação de domínios protéicos em regiões afetadas por deleções e inserções. Esse recurso também foi utilizado para descrever os possíveis produtos codificados por novas ORFs detectadas em alguns isolados geográficos de AgMNPV. As buscas foram realizadas contra o banco de dados "CDD v3.10 – 44354 PSSMs"; com um limiar de *e-value* igual a 0,01; e filtragem de regiões de baixa complexidade.

3.13 ESTIMATIVAS DE DIVERSIDADE GENÉTICA

A diversidade genética dos genes de AgMNPV foi avaliada a partir do programa DnaSP (versão 5) (LIBRADO & ROZAS, 2009). A partir de alinhamentos múltiplos, este programa estima a diversidade genética com base no número de sítios segregantes, utilizando o estimador de Watterson ($\theta = 4Ne.\mu$) (WATTERSON, 1975), onde θ é a medida de diversidade (taxa de mutação populacional), *Ne* é o tamanho efetivo da população, e μ é a taxa de mutação por geração (TAJIMA, 1993). A sequência codificante de cada gene dos 17 isolados foi extraída dos genomas com o auxílio de um *script* PERL (*ripseq.pl*), o qual, a partir das coordenadas genômicas de início e fim das ORFs, extrai as sequências do códon de iniciação ao de terminação. Os alinhamentos múltiplos foram gerados usando *ClustalW* (THOMPSON *et al.*, 1994) e manualmente editados com o programa *Bioedit* (HALL, 1999).

3.14 ORGANIZAÇÃO GENÔMICA E SINTENIA

A organização genômica dos alfabaculovírus do grupo I, entre os quais está incluído o vírus AgMNPV, foi analisada quanto à conservação do ordenamento físico dos genes (sintenia). Para isto, foi utilizada uma abordagem de genômica comparativa que envolveu 13 baculovírus: AgMNPV (DQ813662); AcMNPV (L22858); AnpeNPV (DQ486030); BmNPV (L33180); BomaNPV (FJ882854); CfDefMNPV (AY327402); CfMNPV (AF512031); EppoNPV (AY043265); HycuNPV (AP009046); MaviMNPV (EF125867); OpMNPV (U75930); PlxyMNPV (DQ457003) e RoMNPV (AY145471). Foram gerados arquivos de comparação por meio de buscas por similaridade entre genomas utilizando o BLASTN implementado no site *WebACT* (ABBOTT *et al.*, 2007). O programa *Artemis Comparative Tool* (CARVER et al., 2005) foi empregado para gerar os perfis comparativos e visualizar a sintenia. Seguindo este mesmo procedimento, a organização genômica dos 17 isolados de AgMNPV foi analisada. A identificação precisa das regiões com inserções e deleções, assim como a localização de regiões repetitivas ao longo do genoma (como *hrs*) foram feitas usando matriz de pontos (*Dot Plot*) construídas por meio do programa *Dotter* (SONNHAMMER & DURBIN, 1995).

4 RESULTADOS

4.1 DADOS DO PIROSEQUENCIAMENTO

4.1.1 Fragmentação do DNA genômico

Para o pirosequenciamento das amostras, os DNAs genômicos presentes em cada amostra foram fragmentados em diversos tamanhos, os quais em sua maioria atingiram comprimento de aproximadamente 1000 bp (Figura 14). A partir desse conjunto de fragmentos foram geradas as bibliotecas de DNA, as quais foram seqüenciadas em dois subconjuntos (como apresentado na seção 3.5, de Materiais e Métodos). Após sequenciados, estes fragmentos resultaram em 17 bancos de dados de leituras (detalhes a seguir).

4.1.2 Características do conjunto de dados

Após o tratamento das sequências para remoção dos adaptadores, todas as leituras foram mantidas nas análises posteriores, e os bancos de dados finais possuíam entre 26474 e 78766 leituras (Figura 15). O sequenciamento das amostras gerou leituras dos mais diversos tamanhos, as quais variaram de 30 bp a até mais de 1000 bp de extensão. Levando-se em consideração apenas as sequências que coincidiram com alguma região do isolado 2D, o tamanho médio das leituras nos 17 conjuntos de dados variaram entre 326 e 369 bp (Figura 16). Partindo-se do pressuposto que o número de leituras por região (cobertura local) tenha sido uniforme para todos os *loci* e tendo o genoma de AgMNPV-2D como referência, estimase que cada conjunto de dados possibilitou uma cobertura média que variou entre 60 e 185 vezes (X) (Figura 17).



Figura 14 – Distribuição de tamanhos de fragmentos da biblioteca de DNA de cada isolado. NPV 26 NPV 27 NPV 27

O tamanho médio dos fragmentos girou em torno de 1000 bp. FU = Fluorescence Unit; bp = pares de base; DNA *Ladder* = DNA controle. Fonte: Gráficos cedidos pelo Dr. Meik Dilcher.



Figura 15 – Número total de leituras presentes em cada conjunto de dados.

Figura 16 - Tamanho médio das leituras obtidas para cada amostra.





Figura 17 - Cobertura média proporcionada por cada conjunto de dados.

4.2 LEITURAS NÃO CORRESPONDENTES A AgMNPV-2D

Após o mapeamento dos conjuntos de dados contra o genoma de referência por meio do programa BLASTN, algumas leituras apresentaram correspondência parcial (como ilustrado na Figura 11F) ou ausência total de correspondência com qualquer região de AgMNPV-2D. Esse efeito se dá por dois motivos: (*i*) os isolados geográficos possuem regiões genômicas ausentes no isolado 2D (resultados de inserções e duplicações); ou (*ii*) existem regiões presentes em AgMNPV-2D, as quais não são encontradas nos isolados geográficos. Quando uma leitura cobre uma dessas regiões com marcantes diferenças estruturais, o BLASTN não reporta o resultado da busca em um HSP único para toda a extensão da leitura. Leituras desta natureza, que possuíam mais de 25 bp não coincidentes com o genoma referência, foram classificadas como leituras de "correspondência parcial". Em cada conjunto de dados, a proporção de leituras com essa característica variou entre 1,89 e 3,16% (Figura 18).

De 8 a 11% do total das leituras mostraram-se internas às regiões presentes apenas nos isolados geográficos, e por conseqüência não possuíam correspondência com nenhum *locus* do isolado 2D (Figura 18).





Mesmo frente aos parâmetros rigorosos de identidade entre sequência, quando incluídas nas montagens, grande parte dessas leituras se agrupava em *contigs* específicos, os quais possuíam flancos que correspondiam a regiões conhecidas de AgMNPV.

Não foi realizada quantificação para determinar qual proporção desses subconjuntos de leituras possuíam origem celular/contaminante. No entanto, dado os parâmetros restritos de montagem apresentados na seção 3.8.3, o risco de inclusão de leituras contaminantes na estrutura final dos genomas é improvável, dada a baixa identidade que estas possuem em relação às demais leituras que reconstroem os diferentes haplótipos de origem viral.

4.3 ORGANIZAÇÃO GENÔMICA DOS ALFABACULOVÍRUS DO GRUPO I

4.3.1 Sintenia dos genomas

Antes das montagens dos possíveis genótipos predominantes de cada população, uma análise de genômica comparativa foi realizada com o intuito de avaliar a sintenia dos alfabaculovírus do grupo I. Para isso, foi construído o perfil de sintenia de 13 alfabaculovírus: AnpeNPV; OpMNPV; CfMNPV; HycuNPV; AgMNPV; CfDefMNPV; EppoNPV; AcMNPV; PlxyMNPV; RoMNPV; MaviMNPV; BmNPV; e BomaNPV. Evidenciou-se apenas 4

rearranjos estruturais (inversões), perdas e ganhos de genes auxiliares, assim como a conservação de agrupamentos gênicos (*clusters*) em todos estes vírus (detalhes na seção 4.3.2).

4.3.2 Regiões de Transcrição Sobreposta (ROTs)

Uma característica interessante do processo de transcrição de genes em alguns alfabaculovírus é a síntese de RNAs que se estendem por mais de uma ORF (isto é, em tandem). Regiões específicas dos genomas desses vírus determinam a síntese destas unidades transcricionais, as quais possuem tamanhos variados; inícios de transcrição independentes; e compartilham o mesmo sítio de terminação 3', o qual se localiza em regiões ricas em Timina (JIN & GUARINO, 2000). Em função desse conjunto de transcritos cobrirem regiões em comum, estes foram denominados "transcritos sobrepostos" (LUBBERT & DOERFLER, 1984). Na literatura, regiões responsáveis pela codificação dessa ampla diversidade de transcritos são denominadas "regiões de transcrição sobreposta" (*Region of Overlapping Transcripts* - ROT) (DANFORD et al., 2011).

Trabalhos anteriores envolvendo AcMNPV (FRIESEN & MILLER, 1985; KOOL & VLAK, 1993; MAINPRIZE *et al.*, 1986) e BmNPV (KATSUMA et al., 2011) mapearam diversos transcritos sobrepostos contra o genoma destes alfabaculovírus. Os dados destes mapeamentos apontam a existência de diversas ROTs. Entre todas as possíveis ROTs identificadas, apenas as 22 que são compartilhadas por todos os alfabaculovírus do grupo I foram consideradas para análises subsequentes. Essas ROTs estão presentes ao longo de todos os genomas, e nelas são encontrados até 61 genes (21 *cores* e 40 auxiliares), os quais estão presentes em praticamente todos alfabaculovírus do grupo I. O número de genes que compõem as ROTs varia entre dois e oito (Figura 19 e Tabela 7). As ORFs localizadas em ROTs apresentam especial conservação quanto à ordem gênica, assim como grande compactação da informação genética, que se dá graças à estreita proximidade física das ORFs, podendo inclusive apresentarem-se sobrepostas umas as outras, isto é, com ausência total de espaços intergênicos (mais detalhes na seção 4.5.1).



Labe	ela 7 - ORFs conservadas localizadas	s em Ke	giões de	Transcr	ição Sob	repostas	s (RUTS) de 13 a	altabacu	ovírus (do grupc	o I.		
	Genes localizados em ROTs	MDM	O-MNBV	CENTRU		acMMBV 6	KFS presen	Ites nas KC	VIS de 13 a	Itabaculov DiwMMDV	Irus D _e MNIDV	Manhuim	DesNDV	DomoNDV
ROT		AnpeNPV	OPININPV	CIMINPV	HycuNPV	AgMINPV (TDetMINPV	EppoNPV	ACMINPV	PIXYMINPV	KomnPV	MaviMINPV	BmNPV	BomaNPV
-	$polh \rightarrow 1629 \ capsid^* \rightarrow pk-1$	1 > 3	1 < 3	145 < 1	1 > 3	151 < 1	148 < 1	135 < 1	8 > 10	8 > 10	6 > 8	1 > 3	1 > 3	1 > 3
5	$\underline{lef-l} \rightarrow acl3$	136 > 137	12 < 13	12 < 13	139 > 140	19 > 20	17 > 18	10 < 11	13 < 14	13 < 14	11 < 12	5 < 6	5 < 6	5 < 6
33	$odv-e26 \rightarrow ac17$	133 < 134	15 > 16	14 > 15	135 < 136	16 < 17	14 < 15	13 > 14	16 > 17	16 > 17	14 > 15	8 > 9	8 > 9	8 > 9
4	$ac26 \rightarrow iap-l \rightarrow lef-6$	112 > 114	40 > 42	35 > 37	113 > 115	39 < 41	35 > 37	32 > 34	26 > 28	25 > 27	23 > 25	17 > 19	17 > 18	17 > 18
5	$ac38 \rightarrow lef-11 \rightarrow 39k/pp31$	125 < 127	22 > 24	22 > 24	127 < 129	23 > 25	22 > 24	20 > 22	36 < 38	36 < 38	33 < 35	25 < 27	27 < 29	27 < 29
9	$ac43 \rightarrow ac44$	105 < 106	48 < 49	43 < 44	~~	47 > 48	43 < 44	40 < 41	43 > 44	43 > 44	40 > 41	31 > 32	34 > 35	34 > 35
٢	$\underline{ac53} \rightarrow lef-l0 \rightarrow \underline{vp1054} \rightarrow ac55 \rightarrow ac56$	$95 < 98 \ddagger$	56 > 60	50 > 54	94 < 98	53 > 57	49 > 53	47 > 51	53 > 56	53 > 57	50 > 54	38 > 42	42 > 45	42 > 45
∞	$ac60 \rightarrow ac59$	91 > 92	62 < 63	56 < 57	91 > 92	59 < 60	55 < 56	53 < 54	59 < 60	59 < 60	56 < 57	43 < 44	47 < 48	47 < 48
6	$pif-\overline{6} \rightarrow met$	83 < 84	2	64 > 65	81 < 82	68 > 69	64 > 65	61 > 62	68 > 69	68 > 69	65 > 66	52 > 53	56>57	56 > 57
10	$ac76 \rightarrow ac75 \rightarrow ac74 \rightarrow ac73$	77 > 80	76 < 79	68 < 71	75 > 78	72 < 75	68 < 71	65 < 68	73 < 76	73 < 76	70 < 73	56 < 59	59 < 62	59 < 62
11	$\underline{ac81} \rightarrow \underline{gp41} \rightarrow ac79 \rightarrow \underline{ac78} \rightarrow \underline{vlf-1} \rightarrow ac76 \rightarrow ac75$	72 > 78	78 < 84	70 < 76	70 > 76	74 < 80	70 < 76	67 < 73	75 < 81	75 < 81	72 < 78	58 < 64	61 < 67	61 < 67
12	$tlp20 \rightarrow ac81 \rightarrow gp41 \rightarrow ac79 \rightarrow ac78$	71 > 75	81 < 85	73 < 77	69 > 73	77 < 81	73 < 77	70 < 74	78 < 82	78 < 82	75 < 79	61 < 65	64 < 68	64 < 68
13	$vp39 \rightarrow cg30$	67 > 68	89 < 90	80 < 81	65 > 66	85 < 86	81 < 82	76 < 77	88 < 89	87 < 88	85 < 86	4	71 < 72	71 < 72
14	$p18 \rightarrow odv-e25$	62 < 63	94 > 95	85 > 86	60 < 61	90 > 91	86 > 87	81 > 82	93 > 94	92 > 93	90 > 91	70 > 71	76 > 77	76 > 77
15	$p48 \rightarrow p12 \rightarrow p40 \rightarrow p6.9$	53 > 56	101 < 104	93 < 96	50 > 53	96 < 99	92 < 95	87 < 90	100 < 103	98 < 101	96 < 99	77 < 80	84 < 87	83 < 86
16	$\underline{odv} \cdot \underline{ec43} \rightarrow ac108$	47 > 48	108 < 109	99 < 100	46 > 47	106 < 107	102 < 103	94 < 95	108 < 109	105 < 106	103 < 104	84 < 85	91 < 92	90 < 91
17	$p24 \rightarrow gp16 \rightarrow pp34/pep \rightarrow p25 \rightarrow alk-exo$	25 < 29	127 > 131	122 > 126	23 < 27	125 > 129	124 > 128	114 > 118	129 > 133	125 > 129	122 > 126	98 > 102	106 > 110	105 > 109
18	$p26 \rightarrow p10$	21 < 22	132 > 133	128 > 129	20 < 21	132 > 133	130 > 131	119 > 120	136 > 137	132 > 133	129 > 130	104 > 105	113 > 114	112 > 113
19	$ie-0 \rightarrow p49 \rightarrow odv-e18 \rightarrow odv-e27 \rightarrow ac145$	13 < 17	138 > 142	134 > 138	12 < 16	137 > 141	134 > 138	123 > 127	141 > 145	136 > 140	133 > 137	109 > 113	117 > 121	116 > 120
20	$\underline{odv\text{-}e18} \rightarrow \underline{odv\text{-}e27} \rightarrow acl45 \rightarrow acl46^* \rightarrow ie\text{-}l$	$11 < 15\ ^{\$}$	$140 > 145^{\ddagger}$	136 > 140	10 < 14	139 > 143	136 > 140	125 > 129	143 > 147	138 > 142	135 > 139	111 > 115	119 > 123	118 > 122
21	$\underline{odv} \cdot \underline{e56} \rightarrow ie \cdot l^* \rightarrow acl46$	10 > 12	144 < 146	139 < 141	9 > 11	142 < 144	139 < 141	128 < 130	146 < 148	141 < 143	138 < 140	114 < 116	122 < 124	121 < 123
22	$Ac5 \rightarrow lef-2$	-	6 < 7	3 < 4	146 > 147	3 < 4	3 < 4	4 < 5	5 > 6	5 > 6	3 > 4	125 > 126	134 > 135	132 > 133
Númer	o de acesso GenBank dos genomas analisados:	: AnpeNP	V (DQ486(030); OpN	INPV (U7	5930); CfN	ANPV (AF	7512031);	HycuNPV	(AP00904	6); AgMN	IPV (DQ81	3662); Cfl	DefMNPV
(AY32	7402); EppoNPV (AY043265); AcMNPV (L22	2858); Plx;	/MNPV (D	0457003)	; RoMNPV	/ (AY1454	471); Mavi	MNPV (E	F125867);	BmNPV (L33180);]	BomaNPV	(FJ882854	t). > senso
positiv	o; < senso negativo; Genes sublinhados são gen	tes core. *	Nesta regi	ão genômi	ica, estas se	quências e	stão no se	nso negativ	/o, e não c	odificam p	roteínas; †	· AnpeNPV	não possu	ii um gene
homólc	ogo a ac56; § AnpeNPV possui a hr1 (753 bp) et	ntre os gen	es homólog	gos a <i>ac14</i>	5 e <i>ac14</i> 6; :	OpMNP ¹	V possui a	hrf-1 (236	bp) entre o	s genes ho	mólogos a	ac145 e ac	: <i>146</i> ; ¹ Anp	eNPV não
possui	um gene homólogos a <i>ac5</i> ; ² OpMNPV não poss	sui um gen	e homólogo	0 a <i>met</i> ; ³ I	HycuNPV n	ão possui ı	um gene ho	omólogo a	<i>ac44</i> ; ⁴ Ma	viNPV nãc	nu inssoq o	n gene hom	ólogo a <i>cg</i> .	30.

4.4 ORGANIZAÇÃO GENÔMICA DOS ISOLADOS DE AgMNPV

4.4.1 Colinearidade genômica de AgMNPV

Anteriormente a reconstrução efetiva dos genomas predominantes, as montagens preliminares dos genomas consenso de cada população viral apontaram a ausência de rearranjos estruturais entre os isolados geográficos e o isolado 2D (Figura 20).

Figura 20 - Perfis comparativos entre a estrutura genômica de AgMNPV-2D (barras horizontais superiores) e os genomas consenso dos 17 isolados geográficos (provenientes de montagens com o MIRA, barras horizontais azuis inferiores).



(continua)



(continuação)

(continua)



(conclusão)

O primeiro perfil apresentado compara o genoma de AgMNPV-2D com EppoMNPV (disposto na parte inferior), onde as faixas azuis que interconectam estes genomas indicam uma região que sofreu uma inversão, aqui apresentada apenas em caráter ilustrativo, indicando de que forma este rearranjo é graficamente representado. As faixas verticais vermelhas indicam regiões coincidentes com alta identidade e colinearidade. As regiões com ausência de faixas ("faixas" brancas) podem indicar a falta de identidade entre regiões específicas, ou até a ausência total de determinados *loci*.

Esta etapa da análise visou apenas descartar a existência de inversões e translocações, as quais poderiam dificultar a reconstrução dos genótipos predominantes utilizando-se a abordagem apresentada na seção 3.8.2. A partir da reconstrução dos genótipos predominantes

confirmou-se a ausência de rearranjos estruturais, assim como foi possível identificar de maneira precisa as deleções e inserções de segmentos genômicos entre os 17 isolados.

Adicionalmente, matrizes de pontos (*dotplots*) dos genomas dos 17 isolados em comparação o isolado 2D evidenciaram que a ordem dos genes se manteve inalterada ao longo do curso evolutivo do vírus, conferindo aos genomas um alto grau de colinearidade (Figura 21 a Figura 38). Os resultados de coincidências aleatórias entre regiões genômicas foram ocultados para melhor visualização das reais coincidências consecutivas entre regiões similares. A linha diagonal principal de cada gráfico representa regiões idênticas em ambos os genomas. Desvios na linha diagonal que a desloca para baixo indicam segmentos genômicos presentes em AgMNPV-2D, porém ausentes num isolado geográfico (Figura 21 a Figura 38, setas vermelhas). Já desvios que a deslocam para direita indicam o inverso, isto é, segmentos genômicos presentes num isolado geográfico e ausentes no isolado 2D (Figura 21 a Figura 38, setas verdes).

Em caráter ilustrativo (como na Figura 20), uma análise comparativa entre AgMNPV-2D e EppoMNPV também foi realizada por *dotplot* (Figura 21). Na matriz, é possível observar claramente como inversões (seta roxa) e amplas deleções e inserções (setas vermelhas e verdes, respectivamente) são representadas. Regiões presentes em ambos os genomas, porém com baixa similaridade, aparecem em pontos de desvio para baixo e para direita na linha diagonal (setas amarelas).

Algumas inserções em relação ao AgMNPV-2D podem ser identificadas de maneira clara nos *dot plots*, como inserções na região intergênica entre 39k/pp31 e *ubi*; e em *hr2*. De maneira semelhante, amplas deleções também são facilmente identificadas: a deleção do gene *bro-a*; o truncamento de *Ag21* e *hisP*; e deleções na região da *hr4* (detalhes na seção 4.7).

Outro aspecto importante é a identificação de regiões homólogas (*hrs*) distribuídas ao longo dos genomas (Figura 21 a Figura 38, setas azuis). No *dotplot*, regiões com repetições em tandem duplicadas pelo genoma são representadas como quadriláteros de cor escura localizados dentro e fora da diagonal principal. Estes pontos, quando projetados sobre o eixo horizontal, indicam a posição genômica destas *hrs* nos genótipos dos isolados geográficos. Dentre as nove regiões homólogas identificadas no isolado 2D, sete estão presentes em todos os isolados geográficos. No entanto, duas regiões homólogas aparentemente não são encontradas em alguns isolados, são elas: hr2b (não detectada nos isolados 37 e 40) e hr3a (não detectada nos isolados 30, 32, 33, 36, 39 e 40). Além das *hrs* anotadas no genoma de AgMNPV-2D, outras quatro possíveis regiões homólogas foram identificadas entre as *hrs* 2 e 2a, e entre as *hrs* 4 e 1 de alguns isolados (indicadas por \star).





localizadas em alguns genótipos reconstruídos.



















4.4.2 Características gerais dos genomas de AgMNPV

Os genótipos predominantes entre os isolados de AgMNPV possuem genomas menores, com reduções de tamanho que variaram entre 0,2% (131936 bp) e 1,4% (130368 bp) quando comparados ao isolado 2D. A proporção de Guanina e Citosina (conteúdo GC) dos isolados mostrou-se maior (entre 44,51 e 44,60%) que o previamente observado para o isolado referência (44,50%). Entre os 17 genomas analisados, 143 ORFs previamente descritas para o AgMNPV-2D são compartilhadas em todos os isolados, sem grandes modificações estruturais. Entre os genótipos reconstruídos, as ORFs relativas aos genes bro-c (ORF11), Ag21 (ORF21), hisP (ORF28), Ag64 (ORF64), he65 (ORF101/102) e pe-38 (ORF148/149) apresentaram modificações em relação ao isolado 2D. A partir de mutações nestas ORFs - tais como substituições, deleções e inserções - sete novas ORFs foram geradas graças a mudanças de fase de leitura (levando ao encurtamento ou expansão da região codificante dos genes); ou fusões e divisões de ORFs (detalhes na seção 4.7). Adicionalmente, foram identificadas 29 ORFs não anotadas no genoma do isolado 2D, das quais 23 estão presentes em todos os isolados, incluindo o isolado AgMNPV-2D. Por "ORFs não anotadas" entende-se ORFs não assinaladas, porém presentes no isolado AgMNPV-2D, sendo portanto ORFs de supostos genes sobrepostos, ou até mesmo novas ORFs adquiridas pelos isolados. Ao todo, foram identificadas 179 ORFs, as quais, como mencionado, não necessariamente estão presentes em todos os isolados. Por fim, incluindo todos os ganhos e perda de fragmentos genômicos, a porcentagem de identidade dos isolados geográficos frente ao isolado 2D variou entre 96,20 e 97,95% (Tabela 8).

	-			Conjunto d	le ORFs		_
Isolado	Número de bases (bp)	GC (%)	Similares às ORFs do isolado 2D	Derivadas de modificações em ORFs do isolado 2D	Não anotadas no isolado 2D	Número total de ORFs	Identidade em relação ao isolado 2D
2D	132239	44,50%	152	-	26	178	-
26	131649	44,57%	145	6	26	177	97,93%
27	130663	44,57%	145	5	26	176	96,50%
28	130603	44,53%	144	6	26	176	96,84%
29	130465	44,59%	144	6	26	176	96,20%
30	130764	44,51%	144	6	25	175	97,11%
31	131936	44,56%	146	5	27	178	97,89%
32	131073	44,56%	145	6	26	177	97,60%
33	130956	44,54%	145	5	26	176	97,25%
34	131236	44,54%	145	5	27	177	96,64%
35	131903	44,57%	145	7	27	179	97,95%
36	131250	44,48%	144	6	26	176	97,54%
37	131576	44,49%	149	1	25	175	97,02%

97,12%

96,81%

97,73%

96,63%

97,81%

Tabela 8 – Aspectos gerais dos genomas de isolados do alfabaculovírus AgMNPV-2D.

4.5 REGIÕES CODIFICANTES DOS ISOLADOS DE AgMNPV

4.5.1 Sobreposição de regiões codificantes e as ROTs

44,50%

44,55%

44,60%

44,52%

44,57%

A presença de genes com regiões codificantes sobrepostas mostrou-se um padrão comum não somente entre os AgMNPV, como em todos os alfabaculovírus do grupo I. Entre as 152 ORFs previamente anotados no isolado AgMNPV-2D (OLIVEIRA et al., 2006), 67 estão sobrepostas a ORFs a montante ou a jusante (Figura 39). A extensão das sobreposições varia entre uma única base, como observado entre *lef6* (ORF39) e *iap1* (ORF40); até compartilhamentos mais severos, como o que ocorre entre *iap2* (ORF70) e *Ag71* (ORF71), onde 148 bases codificantes são comuns aos dois genes. Proporcionalmente, o gene *lef10* (ORF54) apresenta a sobreposição mais extensa. Neste gene, 73,5% das bases da região codificante são compartilhados com ORFs de genes a montante (*Ag53*) e a jusante (*vp1054*) (Figura 40).


Figura 39 – Genes de AgMNPV com sobreposições em regiões codificantes.

Das 152 ORFs anotadas em AgMNPV-2D, 67 (apresentadas acima) compartilham bases com ORFs vizinhas.





A) A ORF54 (*lef10*) possui a maior proporção de bases sobrepostas: apenas 63 bases são exclusivas deste gene (indicas pela seta), sendo as outras 171 (73,5%) compartilhadas com os genes vizinhos (*Ag53* e *vp1054*). B) Sobreposições observadas no *locus* de *iap2* (ORF70). Em destaque (seta), está a maior sobreposição observada (em número de bases) entre as ORFs anotadas em AgMNPV-2D.

Esse padrão de sobreposição entre ORFs foi observado em todos os 13 alfabaculovírus analisados na seção 4.3.1. Se comparadas com outras regiões genômicas supostamente não envolvidas com transcrição sobreposta, um grande número de ORFs com sobreposições em AgMNPV localizam-se em ROTs ($\chi^2 = 15,189, p = 0,0001$; d.f. = 1). Neste vírus, 39 (63,9%) dos 61 genes presentes em ROTs possuem algum nível de sobreposição em suas ORFs, enquanto apenas 29 (31,9%) dos 91 genes fora de ROTs possuem tal característica (Figura 41A). Ademais, genes que se localizam em ROTs possuem média de comprimento do espaço intergênico inferior a média observada para genes localizados fora destas regiões (Figura 41B). Os *loci* representados na Figura 40 ilustram o alto grau de compactação da informação genética dessas regiões, como se observa entre as cinco ORFs que compõem a ROT 7 (quadro A), e as duas que formam a ROT 9 (entre *pif-6* e *met*).





A) Grande proporção das ORFs presentes em ROTs não possuem qualquer espaço intergênico, isto é, possuem sobreposições com ORFs a montante ou a jusante. Já ORFs localizadas fora de ROTs, em geral, apresentam-se mais espaçadas umas as outras. B) Em média, ORFs incluídas em ROTs possuem tamanho de espaço intergênico inferior a observada em ORFs fora dessas regiões.

Todas as 22 ROTs identificadas nos 13 alfabaculovírus apresentados na seção 4.3 mantiveram-se conservadas entre os 17 genótipos de AgMNPV reconstruídos (Figura 42). As ROTs aqui apresentadas correspondem apenas às 22 compartilhadas por todos os alfabaculovírus do grupo I, o que não descarta a possibilidade de que outros genes de AgMNPV também sejam transcritos em tandem.





4.5.2 Conjunto de genes presentes entre os isolados de AgMNPV

Tendo em vista que genes com ORFs sobrepostas são comuns entre os alfabaculovírus, as anotações genômicas foram realizadas levando em consideração esse padrão. Todos os genes com até 73,5% de sobreposição e tamanho mínimo de 111 pares de bases foram anotados nos 17 genótipos em estudo. A sobreposição máxima foi determinada com base na ampla sobreposição observada para o gene *lef10* de AgMNPV. O limiar inferior de tamanho foi escolhido com base na menor ORF estabelecida previamente entre os alfabaculovírus sequenciados: a ORF79 de CfDefMNPV (111 bp).

Seguindo este parâmetro, foi possível identificar 27 ORFs de novos supostos genes, com ou sem sobreposições, as quais se localizam nos espaços intergênicos entre as 152 ORFs pré-estabelecidas em AgMNPV-2D. Além destas 179 ORFs, adicionalmente foram anotadas outras 10 novas ORFs (mais detalhes nas seções a seguir). Ao todo foram identificadas 189 ORFs entre os 17 genomas analisados, das quais 165 ORFs são compartilhadas por todos os isolados de AgMNPV, sendo 143 correspondentes a ORFs presentes no isolado 2D (Figura 43). Para mais detalhes, vide o Anexo A - Tabelas Suplementares.

4.6 ORFs DE SUPOSTOS NOVOS GENES DE AgMNPV

Espalhadas por todo o genoma, as 27 novas supostas ORFs anotadas nos isolados de AgMNPV possuem tamanho médio de 174 bp (\pm 71) (Figura 44). Entre estas, 18 possuem sobreposições com outras ORFs de AgMNPV, as quais compartilham de 3 a 70% de suas bases (Figura 45). Sobreposições com regiões homólogas (*hrs*) também foram observadas, porém em apenas 4 ORFs: ORF11c (Figura 44C), ORF21c (Figura 44D), ORF117b e ORF117c (Figura 44K). Entre estas 27 ORFs, 21 estão presentes em todos os isolados de AgMNPV, incluindo o 2D.



Figura 43 – Conjunto total de 189 ORFs presentes entre os isolados de AgMNPV.

Os dois anéis de cor predominantemente vermelha representam o genoma do isolado AgMNPV-2D, em senso (anel externo) e anti-senso (anel interno), e as numerações correspondem aos números da ORFs do isolado referência. As ORFs adicionais, dispostas externa e internamente ao genoma, representam ORFs de supostos novos genes não anotados (porém, em sua maioria presentes) no isolado 2D, dispostas em suas posições relativas no genoma referência. As setas pretas indicam a localização de genes compartilhados por todos os baculovírus.

	(e amare	elo)		-					_				_
Α	Aglb		bro-a		В	Ag10	D			C				aı
		A	g5			Ag9b								
	800	<u>1</u> 600	2400	3	4800	5600	6400	7200	8	<u>6</u> 400	7200	8000	8800	9
nolh	1/m2) m4			ntnl/hm									
poin	Agz [2	_		pcp1/bvp		bro-	·c	<	bro-	-c	Agllc i	f2	
	Ag2b	Ag	4b											
D					Е	Ag31b	br	o-d		F				
Ag21	F-protein		Ag2	23		sod						Ag50b		
,				÷										
16800	17600	18400	19200	2	24000	24800	25600	26400	2	39200	40000	40800	41600	4
	Ag21d		Ag22	b	rp-like				iap:					
	AGZIC										ecm	lef8		
G				b	ы				_			1		_
9	Ag51b		lef	10							Ag	104b f		
	ha22		hrr53	0105			Ag102b	nnl		Act104				
	1		1,000				j			1		1		
42400	43200	44000	44800	4	86400	87200	88000	88800	E	90400	91200	92000	92800	9
			_										_	
		·	Ag52b										Ag106	_
					he65a	he65b							od	v-ec43
J Aq113	b bro-q	Agl16			к				_	L	4g123c			ap 16
	,	pifl					_						p24	
_		_	_	÷	_	Agl18	Ag121	V		-	_	_	_	
97600	98400	99200	100000	L	100800	101600	102400	103200	L	104000	104800	105600	106400	L
pif	3	Agll6c				Ag12				Agl2	23b 54		Ag124b	
		AGII6D			Ag Ag117h	Ag119	AGIZUB	v-trex	lef7					
M	D				N					0				
		2 26	D				Agl:	360		-	Agl47	Agl4 pe-3	8	Ag150
		p26	p10		bro-n		Ag136	.eo j b	p49	♪ 146				
1,00600	110400	111200	112000	Ь	115200	116000	116900	117600	h	25600	126400	1,22200	1128000	1.26
103000	LT0400	LT1200	LT12000	Ľ	112200	TT-0000	TT0000	FT 1000	Ц	5000	T70400	LT2 / 200	H79000	120
	Ag130													
	Ag13	ЮЪ	p74		me	:53								
	Aq	j 131												pl

Figura 44 – Regiões genômicas de localização das 27 supostas ORFs de AgMNPV (em verde e amarelo)

As ORFs representadas nos quadros (de A a O) estão coloridas de acordo com a Figura 43. Na porção superior dos quadros, estão as ORFs em senso, e na inferior, ORFs em antisenso. Com exceção dos quadros B e O, todos os outros são representações de ORFs anotadas no genoma de AgMNPV-2D.



Figura 45 – ORFs de supostos genes de AgMNPV-2D com diferentes níveis de sobreposição em relação a outras ORFs.

4.7 VARIAÇÕES ESTRUTURAIS EM RELAÇÃO AO AgMNPV-2D

4.7.1 A ORF2b de AgMNPV-2D

A ORF2b (Ag2b) foi a única suposta ORF presente apenas em AgMNPV-2D. Esta ORF, que mede 129 bp, encontra-se entre as ORFs 2 (Ag2) e 3 (lef2), com 88 bp (68,2%) de sobreposição a montante (Figura 46A). Em todos os 17 isolados geográficos, o gene Ag2apresenta uma inserção única de adenina na posição 11, a qual tira de fase as 10 primeiras bases desta ORF e ocasiona sua fusão com a ORF2b (Figura 46B). A fusão das regiões codificantes gerou a incorporação dos 18 códons iniciais da ORF2b na ORF2 (Figura 46C). Como conseqüência, os 18 primeiros aminoácidos do peptídeo codificado pela ORF2 dos 17 isolados geográficos se assemelham aos que compõem o hipotético produto codificado pelo suposto gene Ag2b (Figura 46D).

Em análises comparativas, 12 dos 18 primeiros aminoácidos codificados pela *Ag2* dos isolados geográficos (Figura 46D, sombreamento verde) também constituem o peptídeo codificado pela ORF homóloga encontrada nos genomas de alfabaculovírus relacionados, como CfDefMNPV e OpMNPV.



Figura 46 – Polimorfismo estrutural no *locus* de *Ag2*.

A) Posição da suposta ORF2b do isolado 2D. B) Locus gênico do Ag2 dos 17 isolados geográficos. Note a incorporação de parte da ORF2b que passa a compor a ORF2. C) Alinhamento de nucleotídeos entre a ORF2b de AgMNPV-2D e a ORF2 dos 17 isolados geográficos. A presença de uma citosina na posição 11 (sombreamento laranja) da ORF2 ocasiona a fusão desta ORF com a suposta ORF2b. Com a mudança de fase de leitura, o códon de início da ORF2 dos 17 isolados passou a localizar-se 51 bases a montante (sombreamento azul). D) Alinhamento de aminoácidos entre Ag2b de AgMNPV-2D e Ag2 dos 17 isolados geográficos. A mudança de fase ocasionada pela inserção da citosina 11 provavelmente ocasionaria a incorporação de 18 aminoácidos originalmente presentes no possível produto de Ag2b(sombreamento verde), e somente a partir do resíduo 18 o produto de Ag2 dos 17 isolados apresentaria similaridade com o peptídeo codificado pela ORF2 de AgMNPV-2D (sombreamento vermelho).

4.7.2 Alterações nos loci de bro-a, bro-b e bro-c

Os *bro* fazem parte de uma família de genes presentes em múltiplas cópias nos genomas de diversas famílias de vírus de DNA fita dupla (BIDESHI et al., 2003). Entre os baculovírus, a quantidade de genes *bro* por genoma pode variar de uma (em AcMNPV) (AYRES et al., 1994) a 16 cópias (em LdMNPV) (KUZIO et al., 1999), no entanto, existem baculovírus onde nenhum gene *bro* foi encontrado. Entre os alfabaculovírus do grupo I, o isolado AgMNPV-2D é o que possui o maior número de cópias de *bro*, 8 ao todo (de *bro-a* a *bro-h*). Entre os 17 isolados geográficos, o número de genes *bro* variou de 7 a 8 cópias, e importantes alterações foram observadas em ORFs que codificam algumas das cópias deste gene.

Em meio aos 17 genótipos reconstruídos, 12 não possuem a ORF do gene *bro-a*, localizada entre a ORF5 (Ag5) e a ORF7 (*bro-b*) (Figura 47A e B). Entretanto, apesar do genótipo predominante da população do isolado 26 não possuir *bro-a* (Figura 47B), foi possível detectar a presença de um haplótipo contendo a ORF6 entre Ag5 e *bro-b* (Figura 47A), porém em baixa freqüência.

O alinhamento da região intergênica imediatamente a jusante à ORF5 (Figura 47A e B, setas) apontam diferentes pontos de quebra de *bro-a* entre *Ag5* e *bro-b* (Figura 47C). Interessantemente, nos genomas onde *bro-a* está presente, um motivo de sequência relativamente conservado de 39 bp é encontrado nas duas regiões intergênicas (5' e 3') desse gene. Em alguns genomas onde *bro-a* está ausente, apenas uma cópia desse motivo é encontrado, enquanto em outros apenas as últimas 18 bases foram mantidas (Figura 47C e D). Esses motivos de sequências poderiam ter ligação com algum mecanismo relativo ao ganho ou perda desse gene em AgMNPV.



Figura 47 – Polimorfismos no locus de bro-a.

A) Haplótipo observado em genomas que possuem o gene *bro-a*. B) Haplótipo observado em genomas onde *bro-a* está ausente. As retas tracejadas delimitam o *locus* gênico perdido em alguns isolados. As setas pretas apontam para regiões intergênicas a jusante da ORF5. C) Alinhamento da região intergênica entre *Ag5* e *bro-a/bro-b* (regiões indicas por setas em A e B). O sombreamento laranja indica regiões com deleções em relação ao genoma do isolado 2D. As triângulos laterais indicam quais genomas não possuem *bro-a*. D) Motivo composto por 39 bp, encontrado antes e depois da ORF de *bro-a*, onde nucleotídeos em negrito indicam identidade entre sequências. As bases destacadas em verde no quadro C representam o motivo observado imediatamente a jusante da ORF5.

A perda de *bro-a* (ORF5) em alguns isolados aparentemente veio acompanhada de mudanças estruturais em *bro-b* (ORF6). Todos os isolados que sofreram deleções no motivo de 39 bp apresentado anteriormente (8 últimas sequências da Figura 47C) também apresentaram múltiplas deleções e inserções de 21 bases na região codificante de *bro-b* (Figura 48). Os dois primeiros *indels* corromperam o domínio DUF4286 (pfam14114), enquanto o último *indel* se deu na região que codifica DUF3627, porém, aparentemente este não acarretou em alterações severas a ponto de descaracterizar a estrutura do domínio.

AgMNPV-2D	AIKQHEEALRQQQEENKQLVTKLIATFSEHSNAMRQELVKKQDFIERVVAMKDKQIEAKDQQVTRVMTDLNRMYNGFQDTMQKKDEQVTKLVAQVI
NPV26_bro-b	AIKQHEEALRQQQEENKQLVTKLIATFSEHSNAMRQELVKKQDFIERVVAMKDKQIEAKDQQVTRVMTDLNRMYNGFQDTMQKKDEQVTKLVAQVI
▶NPV30_bro-b	AIKQHEEALRQQQEENKQLVTKLIATFSEHSNAMRQELVKKQDFIERVVAMKDKQIEAKDQQVTRVMTDLNRMYNGFQDTMQKKDEQVTKLVAQVI
NPV31 bro-b	AIKQHEEALRQQQEENKQIVTKLIATFSEHSNAMRQELVKKQDFIERVVAMKDKQIEAKDQQVTRVMTDLNRMYNGFQDTMQKKDEQVTKLVAQVI
▶NPV33 bro-b	AIKQHEEALRQQQEENKQIVTKLIATFSEHSNAMRQELVKKQDFIERVVAMKDKQIEAKDQQVTRVMTDLNRMYNGFQDTMQKKDEQVTKLVAQVI
NPV35_bro-b	AIKQHEEALRQQQEENKQIVTKLIATFSEHSNAMRQELVKKQDFIERVVAMKDKQIEAKDQQVTRVMTDLNRMYNGFQDTMQKKDEQVTKLVAQVI
▶NPV38 bro-b	AIKQHEEALRQQQEENKQLVTKLIATFSEHSNAMRQELVKKQDFIERVVAMKDKQIEAKDQQVTRVMTDLNRMYNGFQDTMQKKDEQVTKLVAQVI
NPV40 bro-b	AIKQHEEALRQQQEENKQIVTKLIATFSEHSNAMRQELVKKQDFIERVVAMKDKQIEAKDQQVTRVMTDLNRMYNGFQDTMQKKDEQVTKLVAQVI
▶NPV42_bro-b	AIKQHEEALRQQQEENKQIVTKLIATFSEHSNAMRQELVKKQDFIERVVAMKDKQIEAKDQQVTRVMTDLNRMYNGFQDTMQKKDEQVTKLVAQVI
NPV43_bro-b	AIKQHEEALRQQQEENKQLVTKLIATFSEHSNAMRQELVKKQDFIERVVAMKDKQIEAKDQQVTRVMTDLNRMYNGFQDTMQKKDEQVTKLVAQVI
▶NPV27_bro-b	AIKQ <mark></mark> QQEENKQLINKLVKTFSDHTNTLQ <mark>TALLQKT</mark> QELVKKQEFIERIVATKDKQIEAKDQQVTRVMTDLNRMYNGFQDTMQKKDE <mark>IMQQKDA</mark> QVTELVAKVI
▶NPV28_bro-b	AIKQ <mark>QQEENKQLINKLVKTFSDHTNTLQTALLQKT</mark> QELVKKQEFIERIVATKDKQIEAKDQQVTRVMTDLNRMYNGFQDTMQKKDE <mark>IMQQKDA</mark> QVTELVAKVI
▶NPV29_bro-b	AIKQ <mark></mark> QQEENKQLINKLVKTFSDHTNTLQ <mark>TALLQKT</mark> QELVKKQEFIERIVATKDKQIEAKDQQVTRVMTDLNRMYNGFQDTMQKKDE <mark>IMQQKDA</mark> QVTELVAKVI
▶NPV32_bro-b	AIKQ <mark>QQEENKQLINKLVKTFSDHTNTLQ<mark>TALLQKT</mark>QELVKKQEFIERIVATKDKQIEAKDQQVTRVMTDLNRMYNGFQDTMQKKDE<mark>IMQQKDA</mark>QVTELVAKVI</mark>
▶NPV34_bro-b	AIKQ <mark></mark> QQEENKQLINKLVKTFSDHTNTLQ <mark>TALLQKT</mark> QELVKKQEFIERIVATKDKQIEAKDQQVTRVMTDLNRMYNGFQDTMQKKDE <mark>IMQQKDA</mark> QVTELVAKVI
▶NPV36_bro-b	AIKQQQEENKQLINKLVKTFSDHTNTLQTALLQKTQELVKKQEFIERIVATKDKQIEAKDQQVTRVMTDLNRMYNGFQDTMQKKDEIMQQKDAQVTELVAKVI
▶NPV37_bro-b	AIKQ <mark></mark> QQEENKQLINKLVKTFSDHTNTLQ <mark>TALLQKT</mark> QELVKKQEFIERIVATKDKQIEAKDQQVSRVMTDLNRMYNGFQDTIQKKDAQVTDLVAKII
NPV39 bro-b	A TKOOCEFNKOLTNKLWKTESDHTNTLOTALLOKTOFLWKKOFFTEDTVATKDKOTFAKDOOVTRUMTDLNDMVNGEODTMOKKDFTMOOKDAOVTELWAKUT

Figura 48 - Polimorfismos no locus de bro-b.

Alinhamento de aminoácidos entre os resíduos 130 e 240 de BRO-B. O sombreamento laranja indica regiões de inserção ou deleção de 21 bp (em relação ao isolado 2D).

Outra importante mutação ocorrida em genes *bro* se deu na ORF11, de *bro-c*. No genoma de AgMNPV-2D este gene codifica um peptídeo que possui o típico domínio Bro-N em N-terminal, e um domínio DUF3627 no C-terminal, o qual por vezes é denominado domínio "Bro-C" (BIDESHI et al., 2003) (Figura 49A). Em alguns dos genótipos reconstruídos, a ORF que codifica *bro-c* apresentou-se separada em duas ORFs independentes (11A e 11B) (Figura 49B), as quais foram geradas em função da presença de *indels* na região codificante. Nos isolados 35 e 36, a ausência de uma única base na posição 246 da ORF11 alterou o quadro de leitura dos códons a jusante (Figura 49C) e causou sua terminação precoce no sítio 435 (Figura 49E). Já em NPV32 a fissão se deu em função de uma mutação do tipo "*slipped-strand mispairing*" (LEVINSON & GUTMAN, 1987), a qual ocasionou a duplicação de uma região da ORF11 pode ter dado origem a dois genes independentes, aqui denominados *bro-c1* e *bro-c2*. A partir destes seriam codificados dois peptídeos distintos, cada um portando um dos domínios originalmente codificados por *bro-c* (DUF3627 e Bro-N, respectivamente) (Figura 49B).



Figura 49 - Divisão da ORF11 em duas ORFs independentes.

A) Localização do gene *bro-c* no genoma de AgMNPV-2D. A parte superior ilustra *bro-c* em ORF única, e a inferior apresenta os dois domínios presentes na proteína codificada pelo gene. B) Localização das duas ORFs (11A e 11B) que codificam *bro-c1* e *bro-c2*. A interrupção precoce do quadro de leitura fez com que cada domínio passasse a ser codificado por uma ORF independente. C) Alinhamento de nucleotídeos entre as posições 225 e 265 da ORF11. O sombreamento laranja indica a ausência de uma base que altera o quadro de leitura e causa sua terminação precoce no sítio 435. D) Em NPV32 a divisão em duas ORFs se deu em função de uma mutação que duplicou um fragmento de 25bp (em amarelo e laranja) e levou ao término do quadro de leitura na posição 388 (em vermelho). E) Alinhamento de nucleotídeos entre as posições 400 e 500 da ORF11. O sombreamento vermelho indica o codon de terminação da ORF11A dos isolados 35 e 36, enquanto os azuis destacam o codon de iniciação da ORF11B.

4.7.3 Separação da ORF21

A ORF21 (Ag21) de AgMNPV-2D possui 444 bp (OLIVEIRA et al., 2006), e suas regiões 5' e 3' possuem similaridade, respectivamente, com as ORFs 19 e 20 de CfDefMNPV (OLIVEIRA et al., 2006). Entre as 17 populações analisadas, apenas duas (isolados 36 e 37) possuem o ag21 codificado por uma ORF única. Nas outras 15 populações virias, uma deleção de 10 bp na porção 5' da ORF21 tirou de fase 35 bases, causando a terminação precoce do quadro de leitura e levando a geração de duas ORFs independentes, denominadas ORF21a e ORF21b (Figura 50). Os peptídeos codificados por Ag21a e Ag21b apresentam grande similaridade com as sequências das regiões N e C terminal do peptídeo de Ag21 do isolado 2D (92% e 94% de identidade, respectivamente). Em relação aos produtos das ORFs 18 e 19 de CfDefMNPV, essa similaridade também existe, porém é menor (respectivamente, 89% (ORF21a) e 51% (ORF21b)).



A) ORF21 de AgMNPV-2D. A seta aponta para o ponto no qual ocorreu uma deleção de 10 bp. B) ORF21a e ORF21b (em laranja). Essas ORFs são produto da separação da ORF21 em dois quadros de leitura independentes. C) Alinhamento de nucleotídeos do *locus* de *Ag21*. O sombreamento laranja representa a região ausente em relação a ORF21 do isolado 2D. O sombreamento azul indica o codon de terminação da ORF21a.

4.7.4 Truncamento do gene hisP

A ORF28 (*hisP*) de AgMNPV-2D possui 546 bp e codifica uma proteína semelhante à histidinol-fosfatase (OLIVEIRA et al., 2006) (Figura 51A), a qual possui os motivos "HAD_like" (cd01427) e "Hydrolase_like" (pfam13242) (Figura 51B). Em 8 populações de AgMNPV, o alelo mais freqüente deste gene possui a ORF reduzida a apenas 234 bp (Figura 51C), a qual codifica para um histidinol-fosfatase desprovida do domínio "Hydrolase_like" e provavelmente sem alguns dos resíduos do sítio ativo (Figura 51D).

Nas populações dos isolados 27, 34, 35 e 36, o gene *hisP* está presente nas formas íntegra e truncada, porém o haplótipo do alelo truncado mostrou-se o mais freqüente.



A) Gene *hisP* integro encontrado em AgMNPV-2D e outros 10 genótipos predominantes reconstruídos. B) Domínios da proteína histidinol-fosfatase. C) Versão truncada do gene *hisP* (em laranja) encontrada em 8 isolados de AgMNPV. D) Domínio HAD encontrado no peptídeo codificado pela versão truncada de *hisP*. A região perdida por essa suposta proteína (região sombreada em preto em A, e delimitada por setas em B compreendia o domínio Hydrolase_like e possuía alguns resíduos do sítio ativo da enzima (triângulos vermelhos acima da barra de escala do domínio HAD_like).

4.7.5 Alterações na ORF64

A ORF64 de AgMNPV-2D foi uma das três ORFs identificadas que não possuem homólogas em outros baculovírus. Com 261 bp de extensão, tem capacidade de codificar um produto com 86 aminoácidos (OLIVEIRA et al., 2006) (Figura 52A). Nos 17 genótipos predominantes dos isolados geográficos, apenas dois isolados (32 e 37) possuem esta ORF com características similares a do isolado 2D. Para todos os outros 14 isolados, a ausência de 32 bases na região 5' da ORF64 altera a fase de leitura e faz com que o quadro se feche 230 bases a jusante, no próximo códon de terminação, mais interno a *DNApol* (ORF65). Isso conduz a formação de uma nova ORF, aqui denominada ORF64b, para distingui-la da ORF comumente observada no isolado 2D. Esta nova ORF, que potencialmente codifica um produto com 153 aminoácidos, possui 462 bases, das quais 262 (56,7%) são compartilhadas com a ORF65 (Figura 52B e C). Os 25 primeiros aminoácidos do possível peptídeo codificado pela ORF64b possuem identidade com o produto da ORF64. Apesar da mudança de fase imposta a todos os outros códons a jusante, a duplicação de uma região de 28 bp posicionada entre as ORF64 e 65 mantém a codificação de um motivo de 9 aminoácidos nos possíveis produtos de *Ag64* e *Ag64b* (Figura 52D).



Figura 52 – Alterações estruturais no *locus* de Ag64.

A) Localização da ORF64 dos isolados 2D, 32 e 37. A seta aponta para a região de 32 bp ausente nos outros 14 isolados de AgMNPV. B) Localização da variante ORF64b. C) Alinhamento de nucleotídeos que evidencia a ausência de 32 bp (em laranja) na ORF de 14 isolados. D) Alinhamento de aminoácidos do suposto peptídeo codificado pelas ORFs 64 e 64b. A região em amarelo ilustra a conservação de resíduos até a posição 25. A ausência dos 32 bp que viriam a seguir ocasionou uma mudança de fase que alterou os códons a jusante, podendo levar a codificação de um peptídeo estruturalmente divergente em relação ao normalmente observado para o isolado 2D (sombreamento vermelho, onde o asterisco (*) ao final indica o *stop codon*). A região em verde indica uma região de 9 resíduos compartilhada pelas duas formas de peptídeos.

4.7.6 Fusão das ORFs 101 e 102 de he65

O gene he65 (ORF105) do alfabaculovírus AcMNPV (AYRES et al., 1994) codifica um peptídeo de 553 aminoácidos, com massa molecular de 65 kDa (BECKER & KNEBEL-MORSDORF, 1993). Em AgMNPV-2D, este gene apresenta-se dividido em duas ORFs, 101 (he65a) e 102 (he65b), as quais, respectivamente, correspondem as regiões 3' e 5' do he65 de CfDefMNPV. A ORF 101 codifica um peptídeo de 217 aminoácidos, enquanto a 102 codifica um com 295 (OLIVEIRA et al., 2006). Entre os isolados de AgMNPV, apenas o genótipo predominante dos isolados 37, 39 e 43 possui he65 separada em duas ORFs, como em AgMNPV-2D (Figura 53A). Para os demais genótipos reconstruídos, as regiões 5' e 3' de he65 apresentam-se codificadas numa região codificante única (ORF101a), similar a AcMNPV, a qual pode levar a síntese de uma proteína de 548 aminoácidos (Figura 53B). A fusão entre as duas ORFs ocorreu em função de uma mutação no primeiro sítio do codon de terminação da ORF101 (TAG \rightarrow CAG) (Figura 53C). Esta fusão faz com que a região intergênica que existia entre as duas ORFs seja incorporada à ORF única. Essa região, constituída por 105 bp, é responsável pela codificação de um domínio de ligação a ácido nucléico (DNA ligase A M (pfam01068)) (Figura 53D), o qual compõe um domínio maior com função semelhante (RNA_ligase (pfam09414)), presente na estrutura expandida de HE65 (Figura 53E).



Figura 53 – Variações da organização gênica no locus de he65.

A) Variante observado no isolado 2D e em outros três isolados de AgMNPV. A seta indica o ponto onde ocorreu uma mutação que ocasionou a fusão das duas ORFs. B) Variante de *he65* fusionada. C) Alinhamento de nucleotídeos de uma região de *he65*. O sombreamento laranja indica o codon que sofreu a mutação TAG \rightarrow CAG. O sombreamento azul indica o códon de terminação da ORF102 (à esquerda) e o de iniciação da ORF101 (à direita). O sombreamento cinza refere-se à região intergênica entre as ORFs 101 e 102. D) Domínio codificado pela região intergênica. E) Domínio presente na estrutura expandida de HE65. As setas delimitam a região onde se localiza o domínio presente na região intergênica.

4.7.7 Polimorfismos estruturais de pe-38

O gene *pe38* tem como produto uma proteína de 38 kDa que possui um domínio de ligação a DNA (RING *Finger*, cd00162). Durante as infecções virais esta proteína se localiza no núcleo celular, onde atua estimulando a replicação de DNA (KOOL *et al.*, 1994; KRAPPA *et al.*, 1995). Em AcMNPV e CfDefMNPV a ORF que codifica essa proteína tem tamanho variando entre 954 e 963 bp. Já em AgMNPV-2D, a ORF única de *pe38* encontra-se subdividida em duas ORFs independentes (148 e 149), as quais possuem 396 e 630 bp, respectivamente (OLIVEIRA et al., 2006).

Entre os 17 isolados de AgMNPV, as duas formas de *pe38* foram observadas, entretanto, a forma única mostrou-se predominante em muitas populações. O alinhamento da região intergênica das ORFs 148 e 149 do isolado 2D apontou que a inserção de uma Adenina na posição 341 da ORF148 foi suficiente para causar a terminação precoce desse quadro de leitura 23 bases a jusante. Ademais, bastaria a simples deleção de uma base nessa região para restabelecer a codificação de PE38 por uma única ORF (Figura 54A). Como pode ser observado, este indel localiza-se em região homopolimérica de adeninas.

Outro dado importante sobre esse gene diz respeito ao seu caráter variável. O alinhamento global da região codificante de *pe38* indicou a ocorrência de diversos eventos de inserção e deleção de amplos fragmentos de DNA. O conjunto de mutações sofrido por esse gene gerou pelo menos 4 alelos variantes, os quais conduzem à síntese de 4 proteínas estruturalmente divergentes, aqui denominadas PE38- α , β , γ , δ e ϵ (Figura 54B). A variante PE38- α é encontrada em AgMNPV-2D, sendo os blocos de 1 a 3 codificados pela ORF148, enquanto a ORF149 codifica os demais. PE38- β corresponde ao variante observado em CfDefMNPV. Já as formas γ , δ e ϵ representam os variantes encontrados nos 17 genótipos de AgMNPV reconstruídos, são eles: PE38- γ nos isolados 26, 30, 32, 36, 37 e 38; PE38- δ nos isolados 27, 28, 29, 31, 34, 35, 40, 42 e 43; e PE38- ϵ nos isolados 33 e 39. A união entre os blocos 7 e 9, presentes nos variantes α , γ , e ϵ , corresponde ao *direct repeat 5 (dr5)* do isolado AgMNPV-2D (OLIVEIRA et al., 2006).

	0		5																				
Δ		1									.									1	1		
~		340	3	350	360	3	370	38	30	390		400	41	.0	420		430		440		450		460
	AgMNPV-2D	AAAAAAACT	GTTGGCA	AACGTGCO	CGGCCAC	AAAGCAA	ATTTAC	GGACGO	TGTAAA	AATTCT	tgaaag	CAGATG	CCCCCC	CACGTO	TGATG	ACCTTA	AGTCG	GCTAA	TACTGA	AGCGC	GCCAA	TGCT	GCGCGA
	CfDefMNPV	AAAAAG-CT	GCTGGCA	AACGTGCO	CGGCCAC	TCAGCAG	GTTCAC	GGACGO	CTTTGAA	AGTTCT	TCAAAG	CAGATG	CCCCCC	CATGCO	TGATG	ACCTTO	AATCA	ACCTAA	TACTGA	AGCGC	GCCAAA	TGCT	GCGCGA
	NPV26 pe38	AAAAAA-CT	GTTGGC	AACGTGCO	CGGCCAC	AAAGCAA	ATTTAC	GGACGO	TGTAAA	AATTCT	TGAAAG	CAGATG	CGCGCC	CACGTO	TGAAT	ACCTTZ	AGTCG	GCTAA	TACTGA	AGCGC	GCCAAA	TGCT	GCGCGA
	NPV27 pe38	AAAAAG-CT	GCTGGCG	GACGTGCO	CAGCCAC	CCTGCAG	GTTTAT	GGACGO	TTTAAA	AGTTCT	TCAAAG	TAGATG	CCCCCC	TACGCO	TGATG	ACCTTO	AGTCI	IGCTAA	TACTGA	AGCGC	GCCAAA	TGCT	GCGCGA
	NPV28 pe38	AAAAAG-CT	GCTGGCG	GACGTGCO	CAGCCAC	CCTGCAG	GTTTAT	GGACGO	TTTAAA	AGTTCT	TCAAAG	TAGATG	CGCGCC	TACGCO	TGATG	ACCTTG	AGTCI	IGCTAA	TACTGA	AGCGC	GCCAAA	TGCT	GCGCGA
	NPV29 pe38	AAAAAG-CT	GCTGGCG	GACGTGCO	CAGCCAC	CCTGCAG	GTTTAT	GGACGO	CTTTAAA	AGTTCT	TCAAAG	TAGATG	CGCGCC	TACGCO	TGATG	ACCTTO	AGTCI	IGCTAA	TACTGA	AGCGC	GCCAAA	TGCT	GCGCGA
	NPV30 pe38	AAAAAA-CT	GTTGGCA	AACGTGCO	CGGCCAC	AAAGCAA	ATTTAC	GGACGO	TGTAAA	AATTCT	TGAAAG	CAGATG	CGCGCC	CACGTO	TGAATA	ACCTTA	AGTCG	GCTAA	TACTGA	AGCGC	GCCAAA	TGCT	GCGCGA
	NPV31 pe38	AAAAAG-CT	GCTGGCG	GACGTGCO	CAGCCAC	CCTGCAG	GTTTAT	GGACGO	CTTTAAA	AGTTCT	TCAAAG	TAGATG	CGCGCC	TACGCO	TGATG	ACCTTG	AGTCI	IGCTAA	TACTGA	AGCGC	GCCAAA	TGCT	GCGCGA
	NPV32 pe38	AAAAAA-CT	GTTGGCA	AACGTGCO	CGGCCAC	AAAGCAA	ATTTAC	GGACGO	TGTAAA	AATTCT	TGAAAG	CAGATG	CGCGCC	CACGTO	TGAAT	ACCTTZ	AGTCG	GCTAA	TACTGA	AGCGC	GCCAAA	TGCT	GCGCGA
	NPV33 pe38	AAAAAA-CT	GTTGGCZ	AACGTGCO	CGGCCAC	AAAGCAA	ATTTAC	GGACGO	TGTAAA	AATTCT	TGAAAG	CAGATG	CCCCCC	CACGTO	TGATG	ACCTTR	AGTCG	GCTAA	TACTGA	AGCGC	GCCAAA	TGCT	GCGCGA
	NPV34 pe38	AAAAAG-CT	GCTGGCG	GACGTGCO	CAGCCAC	CCTGCAG	GTTTAT	GGACGO	CTTTAAA	AGTTCT	TCAAAG	TAGATG	CGCGCC	TACGCO	TGATG	ACCTTO	AGTCI	IGCTAA	TACTGA	AGCGC	GCCAAA	TGCT	GCGCGA
	NPV35 pe38	AAAAAG-CT	GCTGGCG	GACGTGCO	CAGCCAC	CCTGCAG	GTTTAT	GGACGO	TTTAAA	AGTTCT	TCAAAG	TAGATG	CGCGCC	TACGCO	TGATG	ACCTTO	AGTCI	IGCTAA	TACTGA	AGCGC	GCCAAA	TGCT	GCGCGA
	NPV36 pe38	AAAAAA-CT	GTTGGCA	AACGTGCO	CGGCCAC	AAAGCAA	ATTTAC	GGACGO	TGTAAA	AATTCT	TGAAAG	CAGATG	CGCGCC	CACGTO	TGATG	ACCTTA	AGTCG	GCTAA	TACTGA	AGCGC	GCCAAA	TGCT	GCGCGA
	NPV37 pe38	AAAAAA-CT	GTTGGCZ	AACGTGCO	CGGCCAC	AAAGCAA	ATTTAC	GGACGO	CTGTAAA	AATTCT	TGAAAG	CAGATG	CGCGCC	CACGTO	TGAAT	ACCTTZ	AGTCG	GCTAA	TACTGA	AGCGC	GCCAAA	TGCT	GCGCGA
	NPV38 pe38	AAAAAA-CT	GTTGGCA	AACGTGCO	CGGCCAC	AAAGCAA	ATTTAC	GGACGO	TGTAAA	AATTCT	TGAAAG	CAGATG	CGCGCC	CACGTO	TGAAT	ACCTTZ	AGTCG	GCTAA	TACTGA	AGCGC	GCCAAA	TGCT	GCGCGA
	NPV39 pe38	AAAAAA-CT	GTTGGCA	AACGTGCO	CGGCCAC	AAAGCAA	ATTTAC	GGACGO	TGTAAA	AATTCT	TGAAAG	CAGATG	CGCGCC	CACGTO	TGATG	ACCTTA	AGTCG	GCTAA	TACTGA	AGCGC	GCCAAA	TGCT	GCGCGA
	NPV40 pe38	AAAAAG-CT	GCTGGCG	GACGTGCO	CAGCCAC	CCTGCAG	GTTTAT	GGACGO	CTTTAAA	AGTTCT	TCAAAG	TAGATG	CGCGCC	TACGCO	TGATG	ACCTTG	AGTCI	IGCTAA	TACTGA	AGCGC	GCCAAA	TGCT	GCGCGA
	NPV42 pe38	AAAAAG-CT	GCTGGCG	GACGTGCO	CAGCCAC	CCTGCAG	GTTTAT	GGACGO	CTTTAAA	AGTTCT	TCAAAG	TAGATG	CGCGCC	TACGCO	TGATG	ACCTTO	AGTCI	IGCTAA	TACTGA	AGCGC	GCCAAA	TGCT	GCGCGA
	NPV43 pe38	AAAAAG-CT	GCTGGCG	GACGTGCO	CAGCCAC	CCTGCAG	GTTTAT	GGACGO	CTTTAAA	AGTTCT	TCAAAG	TAGATG	CGCGCC	TACGCO	TGATG	ACCTTO	AGTCI	IGCTAA	TACTGA	AGCGC	GCCAAA	TGCT	GCGCGA
R							1 1				1 1			1 1	I I	1 1	1 1	1			1 1	1	1 1
-																	_		_		_		
	PE38-α																		\sim				
	BE00 0										_	_	_										
	PE38-β										\sim												
					_																_	_	
	PE38-γ																		\sim				
																	_	_				_	
	PE38-0												\sim										
	DE00 -				_											_				~ ~	_		
	PE38-6																		\sim				
		1		2			3			4		5	6		7		8	9	10	11	12		13

Figura 54 – Variações estruturais em *pe38*.

A) Alinhamento de nucleotídeos do *locus* de *pe38*. O sombreamento laranja indica a ausência de uma Adenina na posição 341 da ORF de *pe38*. O sombreamento em vermelho destaca o codon de terminação da ORF148 (*pe38a*), enquanto o azul indica o codon de início da ORF149 (*pe38b*). B) Variantes estruturais da proteína PE38. Cada retângulo representa um bloco de aminoácidos conservado entre os diferentes tipos de PE-38 (de α a ε). As linhas que os interconectam indicam ausências dos blocos indicados pela numeração na base. Os retângulos fragmentados (em 5, 7 e 9) significam blocos colapsados, presentes apenas parcialmente nos peptídeos. As marcações no topo do quadro representam uma escala de tamanho com intervalos de 10 aminoácidos. A PE-38 α é encontrada em AgMNPV-2D; a β corresponde à PE38 de CfDefMNPV, enquanto γ , δ e ε representam as variantes observadas nos 17 isolados geográficos.

4.8 EVIDÊNCIA DE ORFS ANINHADAS EM BACULOVÍRUS

Além das 152 ORFs anotadas para o isolado 2D, outras 428 ORFs com mais de 100 bases foram identificadas no genoma desse vírus. A partir dos dados de espectrometria de massa (BRACONI; ZANOTTO, em fase de elaboração), e tendo as sequências dos peptídeos codificados por essas ORFs como banco de dados de busca, foi possível identificar uma nova proteína expressa por AgMNPV-2D. Com 46,67% de cobertura, a proteína identificada possui 45 aminoácidos e é codificada por uma ORF localizada internamente (aninhada) à ORF29, do gene fgf (Figura 55A). Esta possível nova ORF (denominada ORF29b) tem 138 bp de extensão, e em função de estar posicionada no sentido oposto (antisenso) à orientação da ORF de fgf, esta codifica um peptídeo que não corresponde a nenhuma região da FGF expressa concomitantemente pelo vírus. Além do isolado 2D, o suposto gene Ag29b é compartilhado e possui 100% de identidade em todos os isolados geográficos de AgMNPV (Figura 55B). Em análises envolvendo outros alfabaculovírus relacionados filogeneticamente, apenas OpMNPV apresentou essa ORF conservada (com 100% de cobertura e 42% de identidade).



A) Localização da ORF29b internamente a região codificante de *fgf*, porém em antisenso. B) Alinhamento de aminoácidos do peptídeo codificado pela ORF29b. Os resíduos destacados em vermelho correspondem aos fragmentos do peptídeo que foram detectados por espectrometria de massa.

4.9 Ag25b, UM GENE ÚNICO DE AgMNPV

Entre as ORFs 25 (*39K/pp31*) e 26 (*ubi*) foi observada uma nova ORF que não possui homólogas em outros baculovírus. Esta ORF, denominada ORF25b, foi encontrada apenas nos isolados 34 e 37, com tamanho variando entre 411 e 420 bp, respectivamente (Figura 56A e B). O peptídeo codificado por essa ORF possui um domínio RING *Finger* (cd00162; pfam13639) na região C-terminal (Figura 56C). RING *Finger* (*Really Interesting New Gene*) é um pequeno domínio com sítio de ligação a zinco, que possui o típico motivo de sequência Cys–X₂–Cys–X_{9–39}–Cys–X_{1–3}–His–X_{2–3}–Cys/His–X₂–Cys–X_{4–48}–Cys–X₂–Cys, e é encontrado em proteínas com funções distintas em diversos organismos (FREEMONT, 2000).



Figura 56 – A ORF25b de AgMNPV. **A**

A) *Locus* onde são encontrados os genes *39K/pp31* e *ubi*, cujas ORFs são separadas por uma região intergênica de 127 bp. B) Variante do mesmo *locus* apresentado em (A), observado nos isolados 34 e 37 (em marrom a ORF25b). C) Domínio RING *Finger* encontrado no C-terminal do possível peptídeo codificados pela ORF25b.

Buscas adicionais no *Conserved Domain Database* (MARCHLER-BAUER et al., 2011) também evidenciaram a presença deste domínio sendo codificado por outros genes de AgMNPV, como: *cg30* (ORF85); *iap-1* (ORF40); *iap-2* (ORF70); *iap-3* (ORF34); *ie-2* (ORF145); e *pe-38* (ORF148). A baixa identidade entre a sequência codificada pela ORF25b e as sequências de RING *Finger* codificadas por estes genes afastam a possibilidade deste domínio ter sido adquirido por duplicação a partir de uma sequência pré-existente no genoma. Por outro lado, por meio de buscas no BLASTP, além da similaridade com domínios semelhantes presentes em diversos organismos, a sequência presente em AgMNPV apresentou significativa similaridade com RING *Finger* de lepidópteros das espécies *Spodoptera exigua* (Noctuidae) e *Papilio xuthus* (Papilionidae) (cobertura, 91%; identidade, 52%; e-value, 2e-19). Assim como *S. exigua*, o lepidóptero *A. gemmatalis* pertence à família Noctuidae, elementos que indicam uma possível origem celular (hospedeiro) dessa nova ORF de AgMNPV (Figura 57).



Figura 57 - Alinhamento de aminoácidos do domínio RING Finger.

As duas primeiras sequências correspondem ao domínio codificado por *Ag25b*. As demais são sequências de RING Finger codificadas por outros genes de AgMNPV, assim como por genes de outros organismos: *Papilio xuthus* (AK402117); *Spodoptera exígua* (GAFU01017521); *Bombyx mori* (AAK57560.1); *Drosophila melanogaster* (2497243); *Picea glauca* (BT103792.1); *Arabidopsis thaliana* (NP_171836.1); *Mus musculus* (AAH06883.1); *Homo sapiens* (1FBV); *Schizosaccharomyces pombe* (CAA20870.1); e *Caenorhabditis elegans* (Q09600). Os resíduos indicados com sombreamento roxo e marrom indicam os aminoácidos que compõem o motivo comumente observado em domínios RING *Finger* (apresentado no topo do alinhamento).

Entre os diversos baculovírus sequenciados, a região presente entre os genes 39K/pp31 e *ubi* é bastante variável. Na maioria dos isolados de AgMNPV, esta região não possui nenhuma ORF, no entanto, em outros baculovírus, o espaço existente entre 39K/pp31 e *ubi* pode apresentar entre 1 e 3 ORFs adicionais, das quais, no entanto, nenhuma apresenta o domínio RING *Finger*.

O alinhamento da região intergênica entre *39K/pp31* e *ubi* dos diferentes isolados de AgMNPV confirma o caráter variável desse *locus*, o qual possui pelo menos três configurações estruturais distintas, aqui denominadas A, B e C (Figura 58).





Cada retângulo representa um bloco de nucleotídeos conservado entre os diferentes isolados de AgMNPV. As linhas que os interconectam indicam ausências de blocos específicos. O retângulo azul fragmentado significa um bloco presente apenas parcialmente. A barra branca interna aos blocos verde, vermelho e amarelo indica a posição do domínio RING Finger, e os círculos nos extremos indicam os códons que codificam os dois resíduos de Cisteína que delimitam cada extremidade do domínio. As marcações no topo do quadro representam uma escala de tamanho com intervalos de 50 bases.

A estrutura "A" (127 bp) é encontrada nos isolados 2D e 32. A estrutura "B" (637 bp) corresponde à região observada nos isolados 34 e 37, e a junção entre os blocos amarelo, vermelho, e parte do verde correspondem a ORF25b (em antisenso). Interessantemente, na estrutura "C" (294 bp), observada nos outros 14 isolados, um novo segmento, com cerca de 135 bp (bloco preto) parece ter sido incorporado a esta região posteriormente, e mesmo que parcialmente, ainda é possível encontrar 90 das 129 bases que codificam o domínio RING *Finger* nessa região. Na estrutura "A" foram encontradas 9 bases que codificam um pequeno fragmento deste domínio.

As evidências apresentadas sugerem que a ORF25b pode ter sido adquirida do hospedeiro por um ancestral de AgMNPV, porém pode ter sido perdida em decorrência de eventos independentes de deleção e inserção ao longo da história evolutiva do vírus.

4.10 DIVERSIDADE GENÉTICA EM AgMNPV

4.10.1 Diversidade em genes core, não-core, e supostos genes

A partir de alinhamentos múltiplos entre as sequências das ORFs apresentadas nas seções 4.5 e 4.6 foi possível avaliar a diversidade genética presente entre os genes de AgMNPV (validados e supostos). A partir dos valores de θ obtidos pelas análises apresentadas na seção 3.13, os dados de diversidade foram agrupados de maneira a constituir 4 categorias de variabilidade: A - Pouco variável (θ entre 0 e 0,01); B - Moderadamente variável (θ entre 0,01 e 0,02); C - Muito variável (θ entre 0,02 e 0,03); e D - Extremamente variável (θ maior que 0,03). Adicionalmente, a categoria A foi dividida em quatro subcategorias: A1 (θ entre 0 e 0,0025); A2 (θ entre 0,0025 e 0,0050); A3 (θ entre 0,0050 e 0,0075); e A4 (θ entre 0,0075 e 0,01).

Tendo em vista que além do isolado 2D, apenas 6 genótipos possuem a ORF6, a diversidade de *bro-a* foi estimada a partir do alinhamento destas 7 sequências. Para os genes que sofreram deleções/inserções que resultaram na divisão da ORF principal em duas, como em *bro-c* (ORF11) e Ag21 (ORF21), o θ foi estimado a partir de um alinhamento múltiplo que incluiu apenas as regiões ainda mantidas em ambos os variantes, isto é, desconsiderando os fragmentos inseridos ou perdidos pela ORF principal (presente no isolado 2D). O mesmo critério foi utilizado para ORFs que sofreram truncamento, como *hisP* (ORF28) e *Ag64* (ORF64). ORFs que nos isolados geográficos se apresentam fundidas em uma única ORF, como as de *he65* (ORFs 101 e 102) e *pe-38* (ORFs 148 e 149), foram tratadas como ORFs independentes, como observado no isolado 2D. Dada a aparente falta de homologia entre os

diferentes blocos de nucleotídeos que compõem a *pe-38b* (ORF149), a diversidade deste gene não foi incluídas nas análises subsequentes.

Das 152 ORFs presentes em AgMNPV-2D e compartilhadas (com ou sem modificações) nos 17 isolados geográficos, 109 (71,7%) agrupam-se nas subcategorias A1 e A2, que representam os genes mais conservados entre os menos variáveis. Apenas 17 genes (11,2%) estão incluídos entre aqueles com variabilidade moderada, elevada ou extremamente elevada.



Figura 59 - Mapa de diversidade genética (θ) de genes de AgMNPV.

Os dois anéis representam um genoma com as ORFs do isolado AgMNPV-2D, em senso positivo (anel externo) e negativo (anel interno). As numerações externas correspondem aos números das ORFs, enquanto as internas dizem respeito às ROTs descritas na seção 4.3.2. As setas pretas indicam a localização de genes compartilhados por todos os baculovírus. As cores no gráfico superior foram estabelecidas de acordo com a escala abaixo, a qual ilustra as categorias e subcategorias de diversidade genética, onde na base estão os intervalos de θ e no topo o número total de genes que se enquadram nas diferentes categorias. Em preto, a ORF149, a qual não foi incluída nas análises.

Entre as ORFs de supostos genes anotados, 20 apresentaram baixa variabilidade, sendo classificadas dentro das categorias A1 e A2. Destas, 4 correspondem àquelas com algum nível de sobreposição com regiões homólogas (*hrs*): ORF11c, ORF21c, ORF117b e ORF117c. Apesar do possível novo gene de AgMNPV (*Ag25b*) possuir considerável variabilidade – observada na comparação entre as sequências dos isolados 34 e 37 – a alta diversidade constatada pelas análises ($\theta = 0,06095$; categoria D) pode estar superestimada, uma vez que esta foi aferida a partir de apenas duas sequências (Tabela 9).

tac	las seguem	o mesmo pa	drão adotad	lo na Fig	gura			
	ORF	Nome	Theta (θ)	Categoria				
	011c* [†]	Agllc	0,00163	A1				
	$021c^{\dagger}$	Ag21c	0,00250	A1				
	029b	Ag29b	0,00000	A1				
	031b	Ag31b	0,00131	A1				
	051b	Ag51b	0,00000	A1				
	052b* [¶]	Ag52b	0,00236	A1				
	$104b^{\dagger}$	Ag104b	0,00216	A1				
	113b	Ag113b	0,00123	A1				
	116b	Ag116b	0,00244	A1				
	116c	Ag116c	0,00000	A1				
	$117c^{\dagger}$	Ag117c	0,00000	A1				
	123c	Ag123c	0,00000	A1				
	130b	Ag130b	0,00195	A1				
	136b	Ag136b	0,00000	A1				
	102b	Ag102b	0,00279	A2				
	$117b^{\dagger}$	Ag117b	0,00349	A2				
	120b	Ag120b	0,00285	A2				
	123b	Ag123b	0,00435	A2				
	124b* [¶]	Ag124b	0,00268	A2				
	147b*	Ag147b	0,00297	A2				
	001b	Ag1b	0,00644	A3				
	021d	Ag21d	0,00525	A3				
	136c	Ag136c	0,00822	A4				
	004b	Ag4b	0,01672	В				
	009b*	Ag9b	0,01312	В				
	022b	Ag22b	0,01062	В				
	$050b^{\dagger}$	Ag50b	0,01581	В				
	025b* [¶]	Ag25b	0,06095	D				

Tabela 9 – Diversidade observada entre as novas ORFs de AgMNPV. Por estar presente apenas no isolado 2D, a diversidade da ORF2b não foi amostrada. As cores aqui representadas seguem o mesmo padrão adotado na Figura 59.

* Supostas ORFs não compartilhadas por todos os isolados.

† Supostas ORFs sobrepostas a regiões homólogas.

¶ Supostas ORFs sem qualquer tipo de sobreposição.

A partir dos valores de θ observadas para as ORFs de AgMNPV, foi possível avaliar como genes de diferentes níveis de conservação estão distribuídos entre as várias categorias de diversidade genética.

Entre os genes *core* (genes ortólogos da família *Baculoviridae*) (GARAVAGLIA *et al.*, 2012; HAYAKAWA *et al.*, 2000; HERNIOU *et al.*, 2001; HERNIOU *et al.*, 2003; JEHLE *et al.*, 2006; ROHRMANN, 2011) e os demais genes, observou-se que a média de diversidade dos 37 *core* (0,0031 ± 0,0028) é significativamente menor que a média obtida para os genes não compartilhados por todos os baculovírus ("não-*cores*") (0,0067 ± 0,0118). A assimetria positiva da curva de distribuição de freqüência dos genes nas diferentes categorias de diversidade apontou uma tendência de agrupamento dos genes *core* na categoria A ($\chi^2 = 3,223$; *p* = 0,0726; d.f. = 1). Por outro lado, a curva de distribuição para os demais genes mostra que estes estão dispersos de maneira mais uniforme entre as quatro categorias de diversidade (Figura 60).

Com o intuito de avaliar o perfil de variabilidade nucleotídica das 29 ORFs dos novos supostos genes de AgMNPV, as análises de diversidade indicaram que a diversidade desses possíveis genes distribuem-se de maneira mais similar aos genes *core*, com valor médio de θ igual a 0,0044 ± 0,0048 (Figura 60). Com exceção das ORFs 9b, 50b, 52b, 104b, 124b, 147b, todas as outras novas ORFs possuem algum nível de sobreposição com ORFs ou *hrs* vizinhas, as quais, indiretamente, podem estar conferindo um maior nível de conservação às supostas ORFs.



Figura 60 – Diversidade genética (θ) de genes *core*, "não-*core*", e supostos genes de AgMNPV.

A) Curvas de distribuições de freqüências dos diferentes grupos de genes dentro das 4 categorias de diversidade genética (apresentadas na base do gráfico, seguindo o mesmo padrão de cores apresentado na Figura 59. B) Média dos valores de θ (eixo y) obtidos para os três grupos de genes e possíveis genes de AgMNPV.

4.10.2 Diversidade por categorias funcionais

Para avaliar a variabilidade genética de genes com base nas funções que esses desempenham durante a infecção viral, estes foram classificados em categorias funcionais com base na principal função reportada para cada gene em revisões recentes sobre o tema (COHEN *et al.*, 2009; GARAVAGLIA *et al.*, 2012; MIELE *et al.*, 2011; ROHRMANN, 2011; VAN OERS & VLAK, 2007). Ao todo, os genes foram incluídos em seis categorias, e genes sem função conhecida foram incluídos numa categoria particular (Figura 61):

i. **Metabolismo de DNA**: genes que participam em diferentes processos relacionados ao metabolismo de DNA, como: ligação a ssDNA; regulação e

síntese de DNA; processamento e proteção contra degradação; recombinação, assim como co-fatores envolvidos nestas atividades;

- *ii.* Capsídeo: genes cujos produtos compõem a estrutura que envolve o DNA e formam os nucleocapsídeos de BV e ODVs;
- *Transcrição*: genes envolvidos em etapas de síntese de RNA; transativação e regulação da transcrição de genes precoces e tardios;
- *iv.* Modulação: conjunto de genes que participam de processos de modulação do funcionamento celular. Entre as principais funções desses genes estão: reorganização do citoesqueleto; inibição da apoptose; e interferências na metamorfose do hospedeiro;
- Auxiliares: genes cujos produtos não são essenciais para a replicação viral mas conferem alguma vantagem a nível celular, de hospedeiro, ou ambiental (O'REILLY, 1997);
- Vírion: genes que codificam proteínas estruturais que participam da formação da formação das partículas virais como um todo, exceto nucleocapsídeo. Entre estas estão proteínas que compõem a matriz e o envelope viral, assim como a estrutura dos corpos de oclusão.



Figura 61 – Mapa gênico por categorias funcionais (OLIVEIRA et al, submetido para publicação)³.

Os dois anéis representam um genoma com as ORFs do isolado AgMNPV-2D, em senso positivo (anel externo) e negativo (anel interno). As numerações externas correspondem aos números das ORFs, enquanto as internas dizem respeito às ROTs descritas na seção 4.3.2. As setas pretas indicam a localização de genes compartilhados por todos os baculovírus.

Tomando-se os valores de diversidade por cada categoria funcional foi possível observar uma baixa média de diversidade genética entre genes implicados na formação de "Capsídeo" e "Vírion". Genes de "Capsídeo" apresentaram média de θ igual à 0,00224 ± 0,0021, enquanto para os de "Vírion" foi de 0,0030 ± 0,0026. De forma mais acentuada do que observado para os genes *core*, a curva de distribuição de freqüência de genes e funções entre as diferentes categorias de diversidade apontou uma significativa concentração de genes de "Capsídeo" ($\chi^2 = 2,869$; p = 0,0903; d.f. = 1) nas subcategorias A1 e A2 (Figura 62). Já a curva de distribuição de genes de função "Vírion" destaca um maior agrupamento destes genes entre A1 e A4. ($\chi^2 = 3,216$; p = 0,0729; d.f. = 1).

² Oliveira, J. V.; Brito, A. F.; Braconi, C. T.; Freire, Caio C. M.; Iamarino, A.; Zanotto, P. M. The modularity and evolutionary constrains in the regulatory network of a baculovirus. Em fase de elaboração

Para as outras categorias funcionais, a distribuição ocorreu de maneira mais homogênea entre as quatro categorias de diversidade, como indicado pelas suas curvas de distribuição, médias e desvios padrões (Figura 62).





A) Curvas de distribuições de freqüências dos diferentes grupos de genes dentro das 4 categorias de diversidade genética (apresentadas na base do gráfico, seguindo o mesmo padrão de cores apresentado na Figura 59. B) Média dos valores de θ (eixo y) obtidos para os seis grupos de genes incluídos nas diferentes categorias funcionais.

5 DISCUSSÃO

5.1 O SEQUENCIAMNETO DE GENOMAS VIRAIS EM LARGA ESCALA

5.1.1 Montagens genômicas num contexto de desequilíbrio de ligação

As populações naturais de baculovírus possuem alta heterogeneidade genética, uma característica decorrente da presença de múltiplos genótipos reunidos em partículas ODV individuais, as quais, por sua vez, encerram-se as dezenas num único corpo de oclusão (FUNK et al., 1997; SIMON et al., 2004). Entretanto, apesar da diversidade genética intrapopulacional, é comum a existência de genótipos dominantes em populações de isolados geográficos coletados em locais e anos diferentes (GELERNTER & FEDERICI, 1990; MUNOZ et al., 1999). A estratégia de montagem adotada neste trabalho visou a reconstrução do provável genótipo predominante de cada população, os quais foram obtidos a partir da junção dos contigs obtidos com maior quantidade de leituras a 97% de similaridade e 150 bp de sobreposição entre si. Desta forma, a união dos haplótipos mais freqüentes de cada região recuperaria aquela que seria a provável estrutura genômica predominante em cada isolado geográfico. Entretanto, sabendo-se que é comum a ocorrência de recombinação entre diferentes genótipos de AgMNPV em situação de co-infecção (CROIZIER & RIBEIRO, 1992), existe a possibilidade dos genótipos reconstruídos não serem os predominantes de cada população, mas sim apenas um entre vários presentes em baixa frequência, uma vez que num contexto de recombinações, as freqüências haplotípicas e genotípicas não são necessariamente iguais, configurando um estado de desequilíbrio de ligação (HEDRICK, 2005) (Figura 63).

As montagens de genótipos individuais num contexto de variabilidade genética aliada a eventos de recombinação mostra-se uma tarefa bastante desafiadora (BEERENWINKEL & ZAGORDI, 2011; ERIKSSON *et al.*, 2008; ZAGORDI *et al.*, 2010). Os dois principais fatores que dificultam as análises de dados de sequenciamento em larga escala de populações geneticamente heterogêneas são: (i) a ocorrência de erros de sequenciamento, principalmente aqueles relacionados a regiões homopoliméricas (HUSE et al., 2007); e (ii) o pequeno comprimento das leituras de sequenciamento (JOJIC *et al.*, 2008).





Os genomas hipotéticos são compostos por 3 haplótipos (A, B e C) que recombinam entre si. Esses haplótipos possuem 3 variantes (1, 2 e 3), os quais estão presentes em diferentes proporções. Neste exemplo, os variantes A1, B1 e C1 são os haplótipos mais freqüentes, no entanto, o genótipo A1B1C1 (primeiro, no topo da imagem: vermelho-verde-azul) não representa o genótipo predominante, e constitui apenas 10% da população total. Por outro lado, o genótipo A2B1C2 (amarelo-verde-roxo) é o genótipo mais freqüente (30%), apesar de nem todos os haplótipos que o compõe serem os mais frequentes na população.

Em função dos pontos mencionados, a recuperação global e precisa da ligação genética entre *locus* fisicamente distantes no genoma torna-se quase impraticável com base nas metodologias de montagem descritas até então (BEERENWINKEL et al., 2012). Nesse sentido, se o objetivo for caracterizar os genótipos individuais dentro de cada população, seria interessante o isolamento prévio dos diferentes variantes genotípicos, para posterior sequenciamento de amostras geneticamente homogêneas, onde uma cobertura final por volta de 25X seria suficiente para este propósito.

Para as 17 populações analisadas, a profundidade de cobertura no sequenciamento variou de 60 a 185X e mostrou-se proporcional a capacidade de detecção de variantes alélicos e genotípicos em cada população. Apesar de não quantificados precisamente, variantes minoritários foram observados para diversos *loci*, fator que confirma o caráter heterogêneo da composição genética intrapopulacional em isolados selvagens de AgMNPV, propriedade já reportada em trabalhos anteriores sobre este vírus (CROIZIER & RIBEIRO, 1992; MARUNIAK, 1989; MARUNIAK *et al.*, 1999).

5.2 ORGANIZAÇÃO GENÔMICA E COMPOSIÇÃO GÊNICA DE AgMNPV

5.2.1 A conservação da ordem gênica em AgMNPV

As análises genômicas das 17 populações de AgMNPV evidenciaram uma alta conservação da ordem gênica dessa espécie viral. Entre vírus pertencentes aos 4 gêneros de *Baculoviridae* (α , β , δ , e γ -*baculovirus*) observa-se uma extensa conservação da ordem gênica (sintenia) (HAYAKAWA et al., 2000; IJKEL et al., 1999; JAKUBOWSKA et al., 2006; WORMLEATON et al., 2003). Por outro lado, deleções e inserções são frequentes em genomas de baculovírus, porém, estas ocorrem em locus específicos, caracterizados pela presença de inúmeras regiões repetitivas (hrs) ou por posicionarem-se próximos a estas (BLISSARD & ROHRMANN, 1990; COCHRAN & FAULKNER, 1983; HAYAKAWA et al., 2000; KUZIO et al., 1999). Análises prévias entre alfabaculovírus relacionados apontam que a distribuição de hrs ao longo dos genomas geralmente se mantém conservada ao longo da história evolutiva desses vírus (ARENDS & JEHLE, 2002), e outra importante função das hrs é atuar em mecanismos de rearranjo genômico e recombinação homóloga, eventos de contribuem para o aumento da heterogeneidade genética em baculovírus (ARENDS & JEHLE, 2002; HAYAKAWA et al., 2000). Para ilustrar, a Figura 19 apresenta inversões observadas nos genomas de 13 alfabaculovírus, as quais ocorreram em regiões flanqueadas por regiões homólogas (AYRES et al., 1994; HYINK et al., 2002; OLIVEIRA et al., 2006).

Além das diversas funções já mencionadas para as regiões repetitivas, uma das mais notáveis é servir como sítio de ligação (*enhancer*) para elementos que regulam a transcrição de genes precoces (CHOI & GUARINO, 1995; RODEMS & FRIESEN, 1995). Para diversos organismos, como bactérias, fungos, protozoários (BORST & GREAVES, 1987), assim como em vírus (WERTZ *et al.*, 1998), sabe-se que rearranjos envolvendo regiões regulatórias podem causar alterações na expressão gênica. A quantidade de *hrs*, assim como a disposição física em relação a diferentes genes apontam a importância destes elementos para a eficiência replicativa ou patogenicidade dos baculovírus (VAN OERS & VLAK, 2007).

Apesar de muito similares em termos de conteúdo e identidade de sequências gênicas, alguns baculovírus possuem um amplo espectro de hospedeiros (HARRISON & BONNING, 1999; HOSTETTER & PUTTLER, 1991), enquanto outros são específicos a poucas ou uma única espécie (GRÖNER, 1986; THIEM & CHENG, 2009). Provavelmente, a capacidade de explorar diferentes ambientes celulares pode ser decorrente tanto da quantidade e distribuição das *hrs* ao longo dos genomas (VAN OERS & VLAK, 2007), como também pode ser reflexo de
rearranjos estruturais proporcionados por estes elementos em regiões genômicas variáveis (COCHRAN *et al.*, 1986; FRIESEN & MILLER, 1986; HARRISON, 2009). O número de *hrs* nos genomas de baculovírus pode variar de três (LANGE & JEHLE, 2003) a 17 (PANG et al., 2001), no entanto, em alguns vírus não foram identificadas quaisquer regiões homólogas (LUQUE *et al.*, 2001; VAN OERS *et al.*, 2005; WILLIS *et al.*, 2005; WORMLEATON *et al.*, 2003). Uma análise mais detalhada das sequências de *hrs* não foi realizada para os 17 genótipos dos isolados de AgMNPV, no entanto, os *dotplots* apontaram a presença de pelo menos 4 possíveis *hrs* adicionais, assim como a ausência de duas *hrs* de AgMNPV-2D em alguns dos isolados (Figura 21 a Figura 38). Essas variações podem interferir (negativa ou positivamente) na regulação de genes precoces durante o ciclo viral desses isolados (CASTRO *et al.*, 1997; COCHRAN *et al.*, 1986; FRIESEN & MILLER, 1986; GUARINO *et al.*, 1986; RODEMS & FRIESEN, 1993).

As interações moleculares que ocorrem entre baculovírus e hospedeiro durante as diversas etapas do ciclo viral dependem da regulação dos níveis de expressão de diferentes fatores virais e celulares (CLEM, 2001; JAKUBOWSKA *et al.*, 2013; OOI & MILLER, 1988; PARK *et al.*, 1996). Independentemente da grande similaridade genética, as diferenças na eficiência replicativa de alguns baculovírus frente a hospedeiros alternativos podem ser decorrentes de incompatibilidades na regulação gênica da interação vírus-hospedeiro (OLIVEIRA et al, em fase de elaboração). Dada a importância da sintenia e das regiões homólogas na regulação da expressão gênica, a conservação ou modificação da estrutura genômica são fatores determinantes na evolução dos baculovírus (HARRISON, 2009). Rearranjos estruturais – como inversões, translocações, e ganhos/perda de genes – podem ser eventos integrantes de um complexo processo de evolução da regulação gênica viral, o qual explicariam a manutenção da colinearidade entre os genomas de AgMNPV.

5.2.2 A compactação da informação genética e as ROTs

Em muitos organismos, as regiões codificantes dos genes dispõem-se separadas umas as outras por longos espaços intergênicos, em geral com mais de 1 Kb, nos quais estão localizados elementos de regulação transcricional, tais como promotores e terminadores (NELSON *et al.*, 2004). Em baculovírus as regiões intergênicas são bastante reduzidas (POSSEE & ROHRMANN, 1997), ou até ausentes, como observado em genes sobrepostos (LU & MILLER, 1994; ZHU *et al.*, 2009). Em vírus de RNA ou DNA, de diversas famílias e comprimentos genômicos, a sobreposição entre regiões codificantes é um fenômeno observado com uma frequência inversamente proporcional ao tamanho do genoma, especialmente em vírus com capsídeo icosaédrico. Restrições físicas impostas pelo tamanho do capsídeo favorecem o encurtamento dos genomas destes vírus, de forma que compactação da informação surge como uma alternativa para a codificação de um maior número de proteínas por um genoma pequeno (CHIRICO *et al.*, 2010; LILLO & KRAKAUER, 2007). Já entre vírus com outras morfologias de capsídeo, como os baculovírus, as sobreposições não parecem existir especificamente para minimizar o tamanho do ácido nucléico (CHIRICO *et al.*, 2010), mas provavelmente atuam em mecanismos regulatórios, tanto a nível transcricional quanto de proteínas (NORMARK et al., 1983).

Estudos sobre a transcrição de genomas compactados, isto é, com espaços intergênicos reduzidos ou ausentes, apontam uma alta frequência de ocorrência de transcrição sobreposta em genomas com estas características (WILLIAMS et al., 2005). Devido às sobreposições, promotores e terminadores da transcrição localizam-se internamente a genes a montante e a jusante, respectivamente. Terminadores "fracos" localizados em genes a jusante podem permitir que a maquinaria de transcrição perpasse mais de um gene, interrompendo a síntese de RNA para além do ponto que determinaria a transcrição de um único gene (CORRADI *et al.*, 2008).

O fenômeno de transcrição sobreposta já foi documentado em baculovírus (FRIESEN & MILLER, 1985; LUBBERT & DOERFLER, 1984; MAINPRIZE *et al.*, 1986), no entanto, a relação entre a compactação da informação genética e a síntese de RNAs em tandem é pouco compreendida. As 22 regiões de transcrição sobreposta (ROTs) identificadas por genômica comparativa entre 13 alfabaculovírus (KATSUMA *et al.*, 2011; KOOL & VLAK, 1993) mostraram-se conservadas não só nestes vírus, como também entre os 17 isolados de AgMNPV, mantendo a mesma orientação (senso positivo ou negativo) e ordenamento gênico, mesmo em situações onde rearranjos estruturais são observados. Interessantemente, a alta frequência de genes sobrepostos ou com espaço intergênico inferior a 50 bp nas ROTs (Figura 64) reitera o padrão observado em estudos com outros organismos que realizam transcrição sobreposta, onde a compactação da informação genética é fator determinante (CORRADI *et al.*, 2008; WILLIAMS *et al.*, 2005).



Os genes em senso positivo e negativo estão representados de maneira linear, em retângulos que delimitam suas respectivas regiões codificantes. As ORFs estão representadas por retângulos cinzas e amarelos, sendo estas últimas ORFs localizadas em ROTs, as quais possivelmente estão envolvidas em transcrições sobrepostas em AgMNPV. Os retângulos roxos indicam regiões homólogas (*hrs*) ou *direct repeats* (*drs*). Os asteriscos (*) evidenciam a posição dos 37 genes *core*. As regiões em rosa indicam espaços intergênicos. De acordo com a legenda, as setas apontam regiões com alta compactação de informação genética, como sobreposições de regiões codificantes ou reduzidos espaços intergênicos. As marcações acima do genoma representam uma escala de tamanho com intervalos de 2200 bases.

Além das 22 ROTs conservadas (OLIVEIRA et al, em fase de elaboração), existem outros *loci* responsáveis pela síntese de RNAs em tandem em BmNPV e AcMNPV (KATSUMA *et al.*, 2011; KOOL & VLAK, 1993), no entanto, apenas aqueles conservados entre 13 alfabaculovírus foram incluídas no presente trabalho. Dada a alta compactação genômica observada não só em AgMNPV (Figura 64), como em diversos outros baculovírus, seria coerente imaginar que diversos outros conjuntos de ORFs sobrepostas ou proximamente posicionadas possam ser sítios de transcrição sobreposta ao longo de todo o genoma, indicando uma alta complexidade transcricional e diversidade de transcritos ao menos entre os alfabaculovírus do grupo I (KATSUMA *et al.*, 2011; KOOL & VLAK, 1993). Se as ROTs de fato determinam a transcrição sobreposta de genes destes vírus, espera-se que aproximadamente 40% dos genes de AgMNPV (isto é, 61, dos quais 20 são *core*) sejam transcritos em tandem (Figura 64), especialmente em estágios tardios do ciclo infectivo do vírus (FRIESEN & MILLER, 1985; KATSUMA *et al.*, 2011; KOOL & VLAK, 1993; LUBBERT & DOERFLER, 1984).

Os múltiplos transcritos em tandem provenientes de ROTs provavelmente atuam principalmente na regulação da expressão de genes virais (FRIESEN & MILLER, 1985; LUBBERT & DOERFLER, 1984; OELLIG *et al.*, 1987), exercendo pouco influência como

mensageiros em processos de tradução, uma vez que apenas a primeira ORF da extremidade 5' é eficientemente traduzida (HAPP *et al.*, 1991; SMITH, 2007). Apesar da proximidade física de genes co-transcritos e de mesma categoria funcional (Figura 61), as evidências disponíveis na literatura afastam a possibilidade das ROTs atuarem como estruturas semelhantes a *operons* (SMITH, 2007), assim, a sequência a jusante da primeira ORF traduzida a partir destes transcritos em tandem provavelmente funcionam como longas UTRs (*Untranslated Region*) (HUGHES, 2006; WILLIAMS et al., 2005), as quais podem ter implicações regulatórias, como em mecanismos de "oclusão de promotores" (ADHYA & GOTTESMAN, 1982) e de RNA interferente (DINGER *et al.*, 2011; HUGHES, 2006).

5.2.3 Diferentes níveis de organização das regiões codificantes

Como discutido anteriormente, os genes de AgMNPV e alfabaculovírus em geral podem se organizar uns em relação aos outros de maneiras variadas, estando separados por amplas ou estreitas regiões intergênicas, ou até mesmo sobrepostos em menor ou maior extensão. Para as anotações do genoma do AcMNPV (AYRES et al., 1994), ORFs com tamanho inferior a 150 bases foram descartadas, parâmetro que proporcionou a identificação de até 337 ORFs, número que se reduziu a 154, a partir do momento em que apenas ORFs contíguas, isto é, sem sobreposições, foram consideradas, com exceções dadas a gene sobrepostos bem caracterizados, como alguns genes *lef (lef-1; 2; 4; 6; e 10)* (POSSEE & ROHRMANN, 1997). A busca por ORFs nos isolados de AgMNPV empregou parâmetros menos restritos, onde ORFs sobrepostas e com menos de 150 bases foram incluídas nas análises preliminares, a qual apontou a presença de 428 ORFs além das 152 previamente anotadas no isolado 2D (OLIVEIRA et al., 2006).

Valendo-se de dados de espectrometria de massa provenientes de um trabalho desenvolvido em nosso grupo (BRACONI; ZANOTTO, em fase de elaboração), a busca por fragmentos de peptídeos associados a partículas virais BV e ODV apontou a expressão de uma proteína codificada por uma ORF localizada internamente a ORF29, do gene *fgf*. A presença de pequenas regiões codificantes internas aos limites de ORFs maiores, e em orientação oposta, é comum entre bactérias e vírus (NORMARK et al., 1983). Em eucariotos esse fenômeno também é observado, porém grande parte das ORFs aninhadas tem localização intrônica, como as encontradas em humanos (LEVINSON et al., 1990) e em organismos modelo, como *Drosophila melanogaster* (CHEN et al., 1987; FURIA et al., 1993; HENIKOFF et al., 1986); e *Mus musculus* (ADELMAN et al., 1987; BOND et al., 1989).

A transcrição bidirecional de genes está descrita para uma série de genes (GIBSON et al., 2005), entre os quais se inclui o *fgf (Fibroblast Growth Factor*), um gene cujo produto é responsável pela regulação de diversos processos metabólicos (BELOV & MOHAMMADI, 2013). O gene aninhado a *fgf*, comumente denominado *gfg* (MURPHY & KNEE, 1994), é expresso em diversos organismos (KIMELMAN & KIRSCHNER, 1989). O possível homólogo presente em AgMNPV também é expresso por meio de RNA antisenso, como apresentado na seção 4.8. Além de AgMNPV, buscas preliminares por TBLASTN também detectaram a presença ORFs semelhantes aninhada a *fgf* de outros baculovírus, como EppoMNPV, OpMNPV, AnpeMNPV e PmSNPV. O suposto produto codificado por *Ag29b* apresentou-se 100% conservado em todos os isolados.

Além da codificação de um peptídeo, outra possível função apontada para os RNAs antisenso de *fgf* é a atuação destes em processos de regulação da expressão de FGF, por intermédio da formação de RNAs dupla fita com os mRNAs de *fgf* (KIMELMAN & KIRSCHNER, 1989).

Com as evidências apresentadas até o momento, qualquer conclusão a cerca da funcionalidade de Ag29b é precipitada, sendo necessários estudos adicionais sobre esse fenômeno. Entretanto, o primeiro registro de uma provável ORF aninhada abre margem para investigações a cerca da existência de um novo nível de organização gênica em baculovírus, o qual ampliaria o grau de complexidade genômica e transcricional destes vírus.

5.3 OS PADRÕES EVOLUTIVOS DE AgMNPV

5.3.1 Origens e implicações evolutivas das sobreposições entre genes

Sequências nucleotídicas específicas do genoma viral podem conter informações genéticas para: (*i*) a codificação de uma ou mais proteínas (CHIRICO *et al.*, 2010; NORMARK *et al.*, 1983); (*ii*) sintetizar mRNAs e ncRNAs (MATTICK, 2011); (*iii*) e definir elementos regulatórios (promotores, iniciadores ou terminadores de transcrição) (TRIFINOV, 1989; XING et al., 2005). Somando-se esses diferentes níveis de organização com a ocorrência de sobreposições entre genes, observa-se que alguns nucleotídeos não somente codificam mais de uma proteína, como também podem conter outros tipos de informações genéticas sobrepostas, o que implica num custo evolutivo adicional, uma vez que mudanças num nível organizacional podem prejudicar o funcionamento em outro (KRAKAUER, 2000).

A principal origem de genes sobrepostos está vinculada ao ganho e perda de códons de iniciação e terminação, eventos que promovem a extensão ou a formação *de novo* de ORFs a partir de genes pré-existentes (CHIRICO *et al.*, 2010), em fenômeno denominado "*overprinting*" (RANCUREL et al., 2009).

Com base no padrão de organização e evolução de genes de AgMNPV, foi possível observar como substituições, inserções e deleções expandem regiões codificantes, podendo ocasionar sobreposições entre genes. Inserções e deleções de bases em números não-múltiplos de três foram frequentes entre os isolados de AgMNPV, e muitos desses eventos — como os ocorridos em Ag2 (*indel* de 1 bp) e Ag64 (*indel* de 32 bp) — ocasionaram a expansão destes genes em sentido 5' e 3', respectivamente. Por outro lado, substituições podem dar origem a novas ORFs a partir de espaços intergênicos e ORFs pré-existentes, como provavelmente ocorreu entre Ag53 e vp1054 para a geração de *lef-10*, um gene presente em apenas alguns baculovírus (GARAVAGLIA *et al.*, 2012; LU & MILLER, 1994). A gradual extensão de ORFs em direção a outras adjacentes, assim como a geração de ORFs *de novo*, resultam em possíveis sobreposições entre regiões codificantes e elementos regulatórios. Como podemos prever, alterações em sequências nucleotídicas específicas podem trazer prejuízos para a regulação transcricional; à estrutura e função dos RNAs transcritos; e, por fim, sobre o dobramento e funcionamento das proteínas traduzidas (Figura 65).



Figura 65 – Prováveis origens e repercussões das sobreposições entre genes de baculovírus.

Os blocos de cor cinza representam uma fita de DNA; os blocos vermelhos indicam ORFs dispostas nos três quadros de leitura possíveis, onde as extremidades coloridas representam, respectivamente, códons de iniciação e terminação. Os círculos (●) representam as regiões promotoras, sendo as setas (→) os sítios de iniciação da transcrição. Os quadrados preenchidos (■) indicam os sítios de terminação; os semi-preenchidos (I) são sítios terminadores menos ativos devido a mutações; enquanto os vazios (
) significam sítios com pouca ou nenhuma atividade terminadora. As linhas sinuosas inferiores denotam os diferentes transcritos codificados, e estão coloridos de acordo com as ORFs imediatamente acima. A) Sobreposição por expansão do quadro de leitura em sentido 3'. Nesta representação, sucessivas substituições, inserções ou deleções alteraram gradualmente o quadro de leitura do gene X. Desta forma, um novo códon de terminação localizado internamente ao gene Y passaria a ser utilizado, causando a sobreposição de regiões codificantes. B) Sobreposição por geração de novo de uma ORF intermediária a duas pré-existentes. O gene Z surgiria a partir de uma ORF delimitada por um códon de iniciação interno ao gene X e outro de terminação interno ao gene Y. Mutações na região codificante do gene Z podem levar a descaracterização do sítio de terminação transcricional do gene X, provocando a expressão de transcritos com terminação a jusante do gene Y (transcrito sobreposto).

Os diversos níveis de organização da informação genética e as sobreposições podem explicar porque subconjuntos de genes são mantidos contíguos e são transcritos em tandem em ROTs. O fato dos transcritos de baculovírus possuírem inícios de transcrição distintos e pontos de terminação coincidentes pode ser reflexo da existência de sítios de terminação

"fracos", localizados internamente e sobrepostos a regiões codificantes (Figura 65) (FRIESEN & MILLER, 1985; GUARINO *et al.*, 1998). Em detrimento das funções regulatórias de terminação, a seleção natural poderia favorecer a manutenção de mutações vantajosas ao funcionamento protéico (KRAKAUER, 2000), assim, a terminação da transcrição se daria em sítios localizados em espaços intergênicos, isto é, livres de sobreposições com regiões codificantes. A nova localização destes sítios de terminação, antecipadas pela perda de capacidade terminadora de sítios a montante, seria essencial para a adequada extensão dos transcritos para além do códon de terminação relativo à proteína a ser expressa. Em termos de tradução da informação genética, a presença de terminadores intermediários poderia levar a síntese de transcritos truncados e sem sentido biológico (Figura 65).

Diante do exposto, pode-se concluir que o alto grau de compactação da informação genética em baculovírus dificulta a separação de genes proximamente dispostos ou sobrepostos, uma vez que o ponto de quebra poderia decompor regiões codificantes, ou romper a ligação entre estas e seus elementos regulatórios, causando efeitos deletérios (CORRADI *et al.*, 2008; HURST *et al.*, 2002). Outro elemento que pode restringir rearranjos é a ocorrência de *splicing* alternativo em baculovírus. A região que compreende as ROTs 19 e 21 contém o gene *ie-1* (Figura 19), até então o único gene de baculovírus que sofre *splicing* (CHISHOLM & HENNER, 1988). Já foi sugerido que a manutenção da colinearidade dessa região se dá em função de possíveis restrições que mantém o posicionamento do sinal de *splicing* no íntron de 4 Kbp que esse gene possui (HAYAKAWA *et al.*, 2000).

Em estudos sobre a sintenia em genomas de outros organismos tem sido discutido se a ordem gênica é mantida por seleção natural, ou é mero efeito de evolução neutra, onde a proximidade física dos genes é conservada unicamente devido à baixa probabilidade de ocorrência de uma separação sem obstrução da codificação e perturbação da regulação de genes adjacentes (HURST *et al.*, 2002; WILLIAMS *et al.*, 2005). Em baculovírus, as evidências aqui apresentadas apontam que as sobreposições e o encurtamento dos espaços intergênicos favorecem a manutenção da ordem gênica de forma a preservar a expressão coordenada de conjuntos específicos de genes, o que poderia afastar a hipótese de evolução por deriva genética.

O caráter essencial da proximidade física de genes para a manutenção da sintenia pode ser claramente observado na região genômica entre as ORFs 62 (posição 49286) e 99 (posição 84378) de AgMNPV-2D. Este *locus* corresponde a 26,5% do genoma de AgMNPV e possui 38 genes, entre os quais estão 21 dos 37 genes *core* (Figura 61). O ordenamento gênico nesta região está extensivamente preservado em diversos alfabaculovírus, e nele se observa grande número de ORFs estreitamente justapostas ou sobrepostas, além de sete ROTs (Figura 64). A composição gênica dessa região difere apenas quanto à presença ou ausência de alguns genes auxiliares, não incluídos em ROTs, como genes *bro* (Figura 19).

Uma vez que o ordenamento e a regulação da expressão gênica em baculovírus estão intimamente associados, os rearranjos estruturais no genoma ficam limitados a regiões flanqueadas por longos espaços intergênicos (HURST *et al.*, 2002) ou por regiões homólogas (ARENDS & JEHLE, 2002; HAYAKAWA *et al.*, 2000).

5.3.2 Fusões e fissões gênicas

Entre os isolados de AgMNPV foram observados duplas de genes adjacentes fundidos, e genes únicos divididos em dois. Análises para inferir as relações evolutivas entre os variantes gênicos combinados ou divididos não foram realizadas, logo, os termos "fusão" ou "fissão" aqui empregados tiveram como ponto de partida os alelos observados no isolado 2D, o que em essência pode não refletir a real polaridade dos eventos. Desta forma, adotando-se como basais os caracteres presentes no genoma referência, foram observados eventos de fissão gênica em *bro-c* e Ag21 (inteiriços em AgMNPV-2D); e outros de fusão gênica em *he65* e *pe-38* (separados em AgMNPV-2D).

Fenômenos de fusões e fissões gênicas ocorrem como forma de, respectivamente, aumentar e diminuir a complexidade de proteínas com arquiteturas multi-domínio (APIC *et al.*, 2001), e em estudos envolvendo Eucariotos, Procariotos e *Archaea*, as fusões gênicas mostraramse mais frequentes que as fissões (KUMMERFELD & TEICHMANN, 2005). A junção de dois genes em um único é mais provável, pois esta envolve a simples perda do códon de terminação do gene a montante, seguida de uma possível (porém não necessária) perda de elementos regulatórios do gene a jusante. Por outro lado, fissões gênicas exigem a aquisição desses fatores genéticos, o que torna estes eventos menos prováveis (KUMMERFELD & TEICHMANN, 2005). Entretanto, como sítios de ligação de fatores de transcrição são compostos por pequenos conjuntos redundantes de nucleotídeos, os quais frequentemente ocorrem ao acaso ao longo do genoma, a associação destes com elementos intensificadores de transcrição (*ehancers*, que atuam a distância), podem tornar exequível a transcrição de genes gerados *de novo* em processos de fissão (TAUTZ & DOMAZET-LOSO, 2011).

Normalmente, novos genes são gerados por fusão ou fissão através de eventos que ocorrem uma única vez em variantes ancestrais, sendo então transmitidos aos seus descendentes (KUMMERFELD & TEICHMANN, 2005). Entretanto, entre os isolados de

AgMNPV foi possível observar evidências da ocorrência de dois eventos distintos responsáveis pela fissão do gene *bro-c*: a deleção de única base (observado em NPV35 e 36) e a duplicação por "*slipped-strand mispairing*" (em NPV32). Aparentemente, para os outros genes, eventos únicos de deleções e substituições de única base foram responsáveis pela união ou separação de ORFs.

A fissão de *bro-c* colocou cada domínio num gene individual. Apesar de fisicamente separados, esses domínios provavelmente ainda se manteriam funcionais e poderiam atuar em conjunto, uma vez que é comum a ocorrência dos domínios Bro-N e DUF3627 ("Bro-C") em proteínas individuais (IYER *et al.*, 2002). Muitas vezes, a geração de novos genes por fusão ocorre entre genes adjacentes e co-regulados (YANAI *et al.*, 2002). Sabendo-se que genes *bro* podem estar envolvidos em mecanismos de recombinação em baculovírus (LI et al., 2002), a separação de domínios individuais se estabeleceriam e eventualmente levariam a formação de proteínas BRO com diversas arquiteturas, por meio de eventos de fusão (Figura 66).



Figura 66 – Diversidade de domínios e proteínas codificadas pelos genes bro de AgMNPV.

A) Conjunto de domínios presentes em proteínas BRO. Ao todo foram observados pelo menos cinco diferentes domínios codificados por AgMNPV. B) Diferentes arquiteturas de domínios em proteínas BRO. O gene *bro-b* codifica dois variantes de proteínas BRO com distintos domínios centrais, como apresentado na Figura 48.

No caso do gene Ag21, sua fissão levou a formação de duas pequenas ORFs, contendo 129 e 156 bp. Haja vista que regiões próximas a hrs são conhecidos hotspots de mutações (BLISSARD & ROHRMANN, 1990; COCHRAN & FAULKNER, 1983; HAYAKAWA *et al.*, 2000; KUZIO *et al.*, 1999), a localização de *Ag21* justaposto a uma região homóloga (*hr1a*) (como *bro-c*, que situa-se ao lado da *hr1*) pode ter contribuído para a ocorrência das mutações que levaram as divisões gênicas.

A arquitetura de domínios do produto de Ag21 não é conhecida, o que dificulta as conclusões a cerca da viabilidade funcional dos dois genes formados pela fissão. Entretanto, o homólogo de Ag21 em CfDefMNPV também encontra-se dividido em duas ORFs (19 e 20) (LAUZON et al., 2005), fator que pode indicar a existência de dois componentes funcionais distintos no produto codificado por este(s) gene(s).

No que tange as fusões entre genes, he65 e pe-38, que estão presentes divididos em duas ORFs no isolado 2D, apresentam-se tanto fundidos quanto separados em diversas das 17 populações de AgMNPV estudadas. As separações destas ORFs em AgMNPV-2D foram decorrentes de substituições ou inserções de única base em pontos específicos próximos ao códon de terminação da ORF a montante. Em ambos os genes, principalmente em he65, a sequência nucleotídica da região imediatamente a montante do ponto de mutação manteve-se conservada, o que implica que provavelmente os elementos regulatórios presentes nestas regiões ainda sejam funcionais. Trabalho anteriores sobre a transcrição e tradução dos genes he65 e pe-38 apontaram a síntese de transcritos com tamanhos diversos, os quais podem ser responsáveis pela codificação de proteínas com pesos moleculares inferiores ao esperado para as ORFs inteiriças (BECKER & KNEBEL-MORSDORF, 1993; KRAPPA & KNEBEL-MORSDORF, 1991; KRAPPA et al., 1995). Neste cenário, seria plausível conceber que tanto gene inteiriço quanto o "gene" localizado em sua porção 3' poderiam ser 0 concomitantemente transcritos e traduzidos de maneira independente por meio dos variantes fundidos destes genes (Figura 67).



Figura 67 – Transcrição e tradução da informação genética num contexto hipotético de fusão ou fissão gênica.

Proteína XY

As representações gráficas seguem o mesmo padrão apresentado na Figura 65. As proteínas são os produtos sintetizados a partir dos transcritos de variados tamanhos, dispostos acima destas. A) Expressão independente dos genes X e Y, os quais estão fisicamente separados por uma região nãocodificante. Neste cenário, as proteínas X e Y são expressas separadamente. B) Expressão do gene fundido XY. A fusão entre os dois genes leva a síntese de um peptídeo de maior extensão (proteína XY). Nessa situação hipotética, uma mutação no códon de terminação inativou a tradução individual da proteína Y, porém não corrompeu o funcionamento dos elementos de regulação transcricional a jusante. Estes possibilitariam a expressão e tradução de transcritos menores, contendo apenas a informação genética para síntese da proteína Y, o que permitiria a tradução de pelo menos duas formas de peptídeos distintas a partir de um mesmo gene (proteínas XY e Y).

A aparente uniformidade observada nas mutações pode afastar a hipótese de ocorrência de eventos múltiplos de fusão e fissão gênica. Entretanto, estudos filogenéticos adicionais — incluindo análises de recombinação entre isolados — são necessários a fim de elucidar melhor a evolução, e possível convergência nestes caracteres.

5.3.3 Perda e ganho de fragmentos genômicos em pe-38

Outro gene que apresentou inserções e deleções foi *pe-38*. Entre os genes que sofreram variações estruturais, *pe-38* foi o mais o que apresentou a maior variabilidade. Este gene codifica uma proteína de 38 kDa, a qual se localiza no núcleo, e outra de apenas 20 kDa, que é encontrada no ambiente citoplasmático durante o estágio tardio da infecção (KRAPPA *et al.*, 1995). PE38 tem ação pró-apoptótica, uma vez que estimula a morte celular por intermédio do produto do gene *ie-1* (PRIKHOD'KO & MILLER, 1999). O produto de 38 kDa possui dois domínios principais (KRAPPA & KNEBEL-MORSDORF, 1991): um RING

Finger em N terminal, que pode intermediar interação entre proteínas ou atuar ligando-se a moléculas de DNA (LOVERING et al., 1993; WU et al., 1996); e um *Leucine Zipper* em C terminal, que promove a dimerização entre proteínas, e também pode ter ação de ligação a DNA (LANDSCHULZ *et al.*, 1988). Este último domínio é caracterizado por motivos compostos por repetições sequências de Leucinas espaçadas por 6 aminoácidos. Na PE38 de AcMNPV, o motivo que forma o *Leucine Zipper* se repete 4 vezes (KRAPPA & KNEBEL-MORSDORF, 1991). Entre os 4 alelos de *pe-38* dos isolados de AgMNPV, a grande diversidade de blocos de aminoácidos apresentada na seção 4.7.7 se deve a várias repetições de motivos de *Leucine Zipper*, os quais possivelmente foram originados por duplicações ou inserções ao longo dos blocos de 4 a 13 (Figura 54). Nestes isolados, os motivos de *Leucine Zipper* são compostos principalmente por leucinas interligadas por resíduos de Ácido Glutâmico, Lisina, Asparagina e Arginina (LEEKNRL), no entanto, outras composições também foram observadas.

No modelo de expressão variada de genes fundidos (como proposto na discussão suportada pela Figura 67B), as diferentes configurações de PE-38 e a possível expressão de dois tipos de proteínas (com 38 e 20 kDa) sugerem que o produto de menor peso molecular possa corresponder ao peptídeo codificado pela região 3' variável do gene. Neste sentido, os diferentes produtos de *pe-38* poderiam atuar em diferentes contextos celulares, onde o peptídeo maior, contendo os dois domínios funcionais, atuaria no núcleo, interagindo com moléculas de proteínas ou ácidos nucléicos, enquanto o menor, contendo as diversas formas de *Leucine Zipper* teria atividade no ambiente citoplasmático, promovendo a homo- ou heterodimerização de proteínas (KRAPPA *et al.*, 1995; PRIKHOD'KO & MILLER, 1999).

5.3.4 Variações haplotípicas envolvendo do gene bro-a

Além de inserções e deleções de fragmentos gênicos, também foi observada a ausência total de uma ORF: apenas cinco isolados geográficos possuíam a ORF6 (*bro-a*) em seus prováveis genótipos dominantes.

A proteína codificada por *bro-a* possui dois domínios: o típico domínio Bro-N em N terminal, com capacidade de ligação a DNA (KANG *et al.*, 1999; ZEMSKOV *et al.*, 2000); e um T5orf172 em C terminal, domínio originalmente encontrado em bacteriófagos T5 (IYER *et al.*, 2002). BRO-A é a única proteína de AgMNPV que possui um domínio T5orf172, a qual é observada em associação com Bro-N em outros baculovírus, assim como também em vírus das famílias *Iridoviridae*, *Poxviridae* e *Phycodnaviridae* (IYER *et al.*, 2002).

Estudos funcionais em BmNPV caracterizaram os genes *bro* deste vírus em níveis de transcrição e tradução (KANG et al., 1999). BmNPV apresenta cinco proteínas BRO: duas com arquitetura de domínio Bro-N + T5orf172 (como na BRO-A de AgMNPV), as quais atuam especificamente no citoplasma; e outras três com arquitetura Bro-N + Bro-C (DUF3627), proteínas comuns entre baculovírus, as quais tem o núcleo celular como sítio de atuação (IYER *et al.*, 2002). A deleção individual da maioria dos genes *bro* de BmNPV não resultou em perdas significativas de infectividade viral, exceto para um destes genes, o qual parece ser essencial para o crescimento viral *in vitro* (KANG et al., 1999).

Devido à atividade de ligação a DNA conferida pelo domínio Bro-N, tem sido proposto que proteínas BRO podem estar envolvidas na regulação de genes virais ou de hospedeiros, ação provavelmente desempenhada por intermédio dos variados domínios em C terminal (IYER *et al.*, 2002). A presença ou ausência de diferentes cópias de genes da família *bro*, assim como a duplicação e diversificação gênica, reforçam a noção de que as proteínas BRO podem atuar determinando o espectro de hospedeiros dos vírus (AFONSO et al., 1999; KANG et al., 1999; LOPEZ FERBER et al., 2001). Ainda não é possível afirmar as conseqüências da ausência de BRO-A em alguns isolados. Esta proteína é a única de AgMNPV que apresenta o domínio T5orf172, responsável por alguma função diferenciada no ambiente citoplasmático (KANG et al., 1999).

Para a população viral NPV26, cujo genótipo predominante possui o haplótipo $Ag5 \rightarrow bro-a \rightarrow bro-b$, a alta cobertura de sequenciamento indicou a presença concomitante do haplótipo $Ag5 \rightarrow bro-b$, encontrado em genomas que não codificam BRO-A. Interações entre vírus e genótipos/espécies de hospedeiros geneticamente distintos poderiam favorecer o aumento ou a diminuição na frequência de um determinado haplótipo ou genótipo viral na população, como esperado nas interações entre variantes virais e hospedeiros resistentes (ASSER-KAISER et al., 2007). A presença dos dois haplótipos numa única população viral poderia indicar a possível ocorrência de flutuações na proporção entre estes haplótipos virais frente a diferentes hospedeiros.

Os mecanismos que originaram ou levaram a ausência do gene *bro-a* nos genomas de AgMNPV é incerta. *bro-a* e *bro-b* são adjacentes, no entanto seus produtos diferem quanto aos domínios em C terminal. As proteínas BRO estão presentes em vírus de eucariotos, fagos, e bactérias, e entre estes organismos existem pelo menos 12 formas distintas de proteínas que possuem Bro-N em sua composição, variedade de configurações que pode ter sido originada pela transferência horizontal e rearranjos entre diversos domínios funcionalmente diversos (IYER *et al.*, 2002). Buscas preliminares no *Conserved Domain Database* (MARCHLER-

BAUER et al., 2011) apontaram que a presença de dois genes adjacentes que codificam proteínas BRO com arquiteturas Bro-N + T5orf172 e Bro-N + Bro-C, um a montante e outro a jusante, respectivamente, é comum em diversos baculovírus, como AgMNPV, BmNPV, HycuNPV, HzSNPV, LdMNPV, além do ascovírus HVAV3e.

Características das regiões que franqueiam bro-a dão indícios sobre a origem do haplótipo bro-a-bro-b e a perda de bro-a durante a história evolutiva de AgMNPV. Como apresentado na Figura 47D, a região de 39 bases a montante do códon de iniciação de bro-a e bro-b possuem alta identidade, característica também observada entre as primeiras 113 bases das ORFs desses genes. A identidade nessa região de 152 bp, que chega a 87,5%, permite supor que o domínio Bro-N da proteína BRO-A pode ter surgido em um ancestral de AgMNPV, pela duplicação da região codificante deste domínio contida em bro-b, o que implicaria na possível aquisição independente do domínio T5orf172 por transferência horizontal (Figura 68A). Por outro lado, o haplótipo bro-a→bro-b pode ter sido incorporado em AgMNPV por meio de um único evento de transferência horizontal (Figura 68B), tendo como possível origem genomas de outros vírus ou procariotos (IYER et al., 2002). A grande similaridade entre as sequências a 5' de bro-a e bro-b também poderia explicar como o gene bro-a pode ter sido perdido. O mau pareamento entre sequências similares não adjacentes pode causar a junção destas em única durante o processo de replicação do DNA, levando a deleção de regiões por "slipped-strand mispairing" (Figura 68C) (LEVINSON & GUTMAN, 1987). Esse conjunto de hipóteses sugere diferentes modos de origem e perda de bro-a em AgMNPV (Figura 68D), entretanto, existem outros fortes indícios que explicam o modo de obtenção e perda deste gene em isolados de AgMNPV, como por exemplo, a aquisição integral e individual deste gene por intermédio de recombinação entre isolados, como será abordado em detalhes na seção 5.3.6. Análises filogenéticas mais rigorosas poderão estabelecer melhor as relações de ancestralidade e esclarecer os eventos por trás da origem e dispersão destes genes nos genomas de AgMNPV.



Figura 68 – Possíveis eventos envolvidos na origem evolutiva e perda do gene bro-a.

A) O gene que codifica Bro-N + T5orf172 (ancestral de *bro-a*) pode ter surgido pela duplicação da região codificante do domínio Bro-N do gene a jusante (evento 1), seguido pela aquisição independente do domínio T5orf172 por transferência horizontal (evento 2). B) O haplótipo *bro-* $a \rightarrow bro-b$ pode ter sido adquirido por meio de um único evento de transferência horizontal em um ancestral de AgMNPV (evento 3). C) Provável deleção de *bro-a* mediada por "slipped-strand mispairing" (evento 4). Ao centro, fita de DNA desnaturada e mal pareada em processo de replicação ($\bullet - - \rightarrow$). N, T e C, representam os domínios Bro-N, T5orf172 e Bro-C, respectivamente. D) Hipotéticas relações filogenéticas entre diferentes haplótipos contendo genes *bro*. Os eventos de duplicação (+), transferência horizontal (+), e deleção (-) indicados ao longo do dendograma refletem aqueles representados em A, B e C. O haplótipo *bro-a → bro-b* de AgMNPV-2D e de outros genótipos pode ter sido originado como os táxons terminais I e III. Os haplótipos que não possuem *bro-a* possivelmente nunca o tiveram, como o táxon V, ou podem ter perdido este alelo, como o táxon II.

5.3.5 Um possível novo gene codificante de RING Finger

O gene Ag25b presente em dois isolados de AgMNPV (34 e 37) possui um domínio RING *Finger*, também codificado por outros genes deste vírus, como: cg30 (PASSARELLI & MILLER, 1994); *iap-1*; *iap-2*; *iap-3* (CLEM & MILLER, 1994); *ie-2* (PASSARELLI & MILLER, 1993); e *pe-38* (KRAPPA *et al.*, 1995). O domínio codificado por Ag25b é do tipo C₃H₂C₃ (RING-H2 *Finger*), variante que possui uma Histidina no lugar da quarta Cisteína do motivo que o caracteriza (FREEMONT, 2000; SAURIN et al., 1996). Nenhum dos RING *Finger* observados nas proteínas de baculovírus é do tipo RING-H2, mas sim RING-HC (C₃HC₄) (Figura 57). Proteínas que apresentam qualquer destes variantes pertencem a diversas categorias funcionais (FREEMONT, 2000), no entanto, estudos têm apontado que tais proteínas – em especial as que possuem RING-H2 – podem participar diretamente em mecanismos de ubiquitinação, atuando como ligases proteínaubiquitina (E3) (JOAZEIRO & WEISSMAN, 2000; LEVERSON *et al.*, 2000; LORICK *et al.*, 1999). Em baculovírus, três proteínas com RING *Finger* parecem atuar como E3, são elas: IAP2, IE2, e PE38 (IMAI et al., 2003). Quase todos os alfa- e betabaculovírus possuem outro componente central da ubiquitinação: a proteína ubiquitina, codificada pelo gene *ubi* (COHEN et al., 2009; GUARINO, 1990; O'REILLY, 1997). Interessantemente, o novo gene encontrado em AgMNPV está localizado a montante do gene *ubi*, e em mesma orientação. Esta proximidade física que pode sugerir uma possível co-regulação destes dois fatores implicados em mecanismos de degradação protéica proteassomo-dependente (O'REILLY, 1997) (Figura 69).





A cascata de reações que conduz a degradação de proteínas pelo proteassomo 26S envolve: ubiquitinas livres (U); enzimas de ativação de ubiquitinas (E1); enzimas de conjugação de ubiquitinas (E2); e ligases proteína-ubiquitina (E3). A proteína codificada por Ag25b provavelmente atua como uma E3, intermediando a interação entre a proteína-alvo (A) e a enzima E2, que por sua vez realiza a ubiquitinação das proteínas-alvo.

Análises preliminares sugerem que este gene pode ter estado presente em ancestrais dos isolados de AgMNPV, uma vez que em todos os genótipos reconstruídos foram encontrados fragmentos do domínio RING *Finger* na região intergênica entre *39K/pp31* e *ubi*, sendo este mantido integralmente em apenas dois isolados (Figura 58).

Para grandes vírus de DNA, como os baculovírus, tem sido sugerido que seus longos genomas são resultado de diversos eventos de transferência horizontal de genes entre vírus e eucariotos; recombinações entre diferentes vírus; e duplicações gênicas seguidas por divergências (HUGHES & FRIEDMAN, 2003; SHACKELTON & HOLMES, 2004).

Os resultados das análises aqui apresentadas são insuficientes para se concluir sobre a origem deste gene. Entretanto, o domínio RING Finger codificado por este gene possui alta identidade com RING-H2 presentes em lepidópteros, além de divergir consideravelmente dos RING-HC observados em outras proteínas de AgMNPV, observações que afastam a possibilidade de origem por duplicação e reforçam a hipótese de origem celular.

5.3.6 Alterações em hotspots de recombinação homóloga

Estudos anteriores envolvendo diferentes isolados de AgMNPV reportaram *loci* variáveis, muitos dos quais estão envolvidos em recombinações entre genomas destes vírus (Figura 70) (CROIZIER & RIBEIRO, 1992; MARUNIAK *et al.*, 1999).





A linha horizontal representa um genoma de AgMNPV, tendo com ponto de início o sítio de restrição que separa os fragmentos *HindIII*-M e G. Os círculos (•) ao longo do genoma indicam ou delimitam regiões variáveis segundo (MARUNIAK et al., 1999). As letras gregas, de α a ε , apontam para os sítios de restrição das enzimas *EcoRI* e *BglII*, os quais segregavam de maneira independente entre genomas após situações de coinfecção, o que indicava a ocorrência de recombinações homólogas, de acordo com (CROIZIER & RIBEIRO, 1992). Setas amarelas apontam a posição de regiões homólogas (*hr*), enquanto as azuis e vermelhas representam genes em senso positivo e negativo, respectivamente.

Eventos de recombinação entre diferentes genótipos de baculovírus podem servir como mecanismos de rápida variação genética, sem passar pelos possíveis efeitos deletérios gerados por altas taxas de mutação (HARRISON, 2009). Alterações no material genético, como inserções, deleções e substituições ocorrem com grande frequência em regiões próximas a *hrs*, e estas mutações contribuem para a geração de novos genótipos em populações selvagens de baculovírus (MUNOZ et al., 1999). Muitas das variações observadas entre os isolados geográficos de AgMNPV coincidem com regiões previamente indicadas como variáveis e *hotspots* de recombinação (CROIZIER & RIBEIRO, 1992; MARUNIAK *et al.*, 1999).

Algumas das divergências estruturais observadas nos isolados geográficos de AgMNPV correspondem a alterações previamente descritas em perfis de restrição de isolados selvagens, em regiões genômicas específicas e até então funcionalmente desconhecidas. Com o sequenciamento de AgMNPV-2D e o resequenciamento de isolados geográficos deste vírus, foi possível identificar em detalhes algumas das regiões afetadas por mutações e recombinações. Entre estes polimorfismos estão os observados em *bro-a*; *bro-c*; *Ag21*; *hisP*; *he65*; *hr4*; *ie-2* e *pe-38*.

No *locus* de *bro-a* é encontrado um sítio de restrição em 4.60 m.u. (*map units*) que delimita os fragmentos G e A gerados por *EcoRI* (JOHNSON & MARUNIAK, 1989). Este sítio (denominado marcador α) juntamente com outros 6 sítios (de β a ϵ) foram utilizados como marcadores de recombinação por (CROIZIER & RIBEIRO, 1992) (Figura 70). O marcador α foi escolhido baseado no fato de alguns genótipos de AgMNPV-B87 não possuírem este sítio, diferentemente de outros que o tinham, o que gerava diferentes perfis de restrição. Os outros marcadores também foram definidos baseados nos diferentes padrões de presença e ausência, tanto no isolado B87, quanto no F3. Os diversos variantes desses isolados foram colocados em contextos de coinfecção a fim de avaliar a ocorrência de recombinações, as quais seriam detectadas com base nas variações dos padrões de presença e ausência de marcadores entre os variantes parentais e seus descendentes.

Buscas por sítios de restrição de *EcoRI* em AgMNPV-2D permitem concluir que a ausência do marcador α em alguns variantes de AgMNPV decorreu da falta do gene *bro-a* nestes vírus. Os diferentes padrões de ganho e perda deste marcador indicam a alta frequência de recombinações neste *locus* gênico, eventos que explicam um importante modo de ganho e perda de genes *bro*, tanto entre variantes de uma mesma linhagem de baculovírus, como também entre outros organismos que possuem membros desta família multigênica.

Com base nessa evidência, outras regiões genômicas de AgMNPV também variariam por recombinação, são estas os *loci* de Ag57; Ag58; he65; a região que compreende de p74 a Ag142; e o *locus* de *ie-2*. Nada afasta a possibilidade de outras regiões gênicas também estarem envolvidas em recombinações homólogas, uma vez que os marcadores utilizados

estavam sustentados apenas na presença ou ausência de curtas sequências de sítios de restrição em seis regiões pontuais dos genomas.

Outras regiões variáveis, porém sem evidências de recombinação com base no trabalho de (CROIZIER & RIBEIRO, 1992), puderam ser avaliadas em detalhes baseado nos achados de (MARUNIAK et al., 1999). Neste estudo, variantes naturais do isolado AgMNPV-85 – genótipo presente na formulação comercial utilizada por anos consecutivos em cultivos de soja (MOSCARDI & SOSA-GOMEZ, 1992) – e AgMNPV-79 foram analisados por perfis de restrição. As regiões variáveis encontradas nestes vírus estão identificadas por círculos (•) na Figura 70, e foram detalhadamente avaliadas com base nos polimorfismos dos isolados geográficos de AgMNPV, os quais possivelmente possuem íntima relação filogenética com o isolado 85.

O ponto 0 (zero) m.u. do genoma de AgMNPV-2D marca o início do mapa físico, e coincide com o sítio de restrição que separa os fragmentos M e G gerados pela enzima *HindIII*. O estudo de (MARUNIAK et al., 1999) evidenciou que nenhum variante do isolado AgMNPV-85 possuia o sítio de restrição de *HindIII* que incide nesta região, que corresponde a porção 3' do gene *pe-38*, um dos genes estruturalmente mais variáveis de AgMNPV (Figura 54 e Figura 59). Com base nas análises apresentadas na seção 4.7.7, foi identificado que o sítio AAGCTT de *HindIII* está mutado para AAGCCT nos genes que codificam a proteína PE38- δ (delta). Esse dado sugere que o alelo δ de *pe-38* provavelmente esteve presente entre os genótipos derivados de AgMNPV-85 (Figura 54). Ademais, este sítio de restrição esteve presente entre todos os variantes de AgMNPV-79, o que pode indicar a presença de algum dos outros alelos de *pe-38* entre os diferentes genótipos deste isolado.

De maneira notável, a região entre os marcadores $\varphi \in \alpha$ (de 96.32 a 4.60), localizada entre a *hr4* e a *hr1* (Figura 70), já havia sido identificada como uma das regiões de mais intensa ocorrência de substituições, inserções ou deleções que corrompem sítios de restrição em AgMNPV (CROIZIER & RIBEIRO, 1992; MARUNIAK, 1989; MARUNIAK *et al.*, 1999). Os dados aqui apresentados sobre diversidade genética dos isolados selvagens de AgMNPV possibilitaram identificar quais genes estiveram implicados nas alterações estruturais previamente reportadas para essa região. Como apresentado na Figura 59, a região entre essas duas regiões homólogas, que se estende da ORF135 a ORF11, concentra a maioria dos genes mais variáveis de AgMNPV, como será discutido em mais detalhes na seção 5.4.

No ponto 5.50 m.u., sete variantes de AgMNPV-85 não possuíam o sítio de restrição *HindIII* que marcava a divisão entre os fragmentos G e J (MARUNIAK et al., 1999). Entre os isolados geográficos foi identificado que este sítio encontra-se na região 3' de *bro-b*, local de extensas substituições, inserções e deleções. Ao todo, oito isolados geográficos não possuíam este sítio, são eles: NPV27, 28, 29, 32, 34, 36, 37 e 39. Para estas amostras foi observada ligação genética entre o *indels* em *bro-b* e a ausência de *bro-a* (Figura 47C e Figura 48A), o que permite supor que os sete variantes naturais de AgMNPV-85 sem o sítio de restrição *HindIII* apresentavam as mesmas variações em *bro-b*, e provavelmente não contavam com o gene *bro-a*.

Entre os pontos 2.43 e 9.20 m.u. foram observadas perdas de 350 bp e inserções de aproximadamente 1100 bp em variantes de AgMNPV-85 (MARUNIAK et al., 1999). Esta região corresponde ao *locus* que se estende da ORF1 (*polh*) a ORF12 (*pif-2*). As evidências são inconclusivas, no entanto, dado que neste locus foi observada a ausência de bro-a, e variações estruturais em *bro-c*, seria possível sugerir a ocorrência de algum evento de deleção ou duplicação de genes bro nesta região. O variante fusionado de bro-c (ORF11) possui 1032 bp, enquanto sua versão fissionada bro-c2 (ORF11b, codificante de um domínio Bro-N isolado) possui entre 393 e 411 bp (Figura 49), medidas que, respectivamente, poderiam corresponder a uma duplicação da ORF11 ou uma deleção da ORF11b. Por outro, os *locus* entre polh e pif-2 também inclui a hr1 (548 bp) (OLIVEIRA et al., 2006), a qual pode ter sofrido inserções e deleções. No entanto, esse tipo de polimorfismo nesta hr não foi detectado entre os isolados geográficos. Já em relação à hrla, localizada entre os pontos 14.70 e 15.69 m.u., foi possível confirmar a ausência de 49 bp entre os isolados geográficos, variação já reportada em genomas de seis variantes de AgMNPV-85 (MARUNIAK et al., 1999). Nesta mesma região também se encontra o gene Ag21, o qual na maioria dos isolados sofreu uma aparente deleção de 10 bp (Figura 50).

Outra importante ausência de sítio de restrição ocorreu no *locus* do gene *hisP*, que como observado no presente trabalho, possui dois alelos distintos (Figura 51). O variante número 18 de AgMNPV-85 aparentemente codifica o variante truncado de *hisP*, uma vez que este não possui especificamente o sítio de restrição de *PstI* encontrado internamente a região perdida pelo alelo truncado. Em 5 isolados geográficos de AgMNPV (27, 29, 34, 35 e 36), ambos os variantes de *hisP* foram encontrados, sendo o alelo truncado o mais frequente em duas populações (29 e 36). Seria possível supor que a região perdida deste gene tenha a função de codificar alguma interface de interação entre HISP e algum fator de defesa celular. Dessa maneira, os dois variantes alélicos poderiam resultar em diferentes respostas nas interações vírus-hospedeiro, podendo ocasionar aumentos ou reduções na capacidade replicativa (*fitness*) viral frente a hospedeiros geneticamente heterogêneos (DAUGHERTY & MALIK, 2012). Esse fator poderia favorecer a manutenção dos dois variantes de *hisP* nas

populações virais e a conversão genética (por recombinação) entre diferentes genótipos de AgMNPV (CROIZIER & RIBEIRO, 1992).

No *locus* delimitado pelos pontos 15.69 e 19.66 m.u., nenhum indício da presença do possível novo gene de AgMNPV (*Ag25b*) foi observado a partir das análises de restrição do DNA de variantes dos isolados 79 e 85 (MARUNIAK et al., 1999). Nestes variantes foi registrada a ocorrência de apenas uma deleção de 100 bp, a qual não possui semelhança com nenhum padrão de deleção observado neste *locus* dos isolados geográficos. Esse dado pode reforçar a suspeita da presença deste gene nos ancestrais de AgMNPV, seguida de pelo menos dois eventos de deleção gênica independentes, como brevemente discutido na seção 5.3.5.

A região genômica onde se encontra a *hr4* de AgMNPV já havia sido reportada como uma das regiões mais variáveis desse vírus (GARCIA-MARUNIAK *et al.*, 1996; MARUNIAK *et al.*, 1999). Este *locus*, correspondente ao fragmento T de PstI (88.58 a 90.13 m.u.) esteve entre os que mais apresentaram variações estruturais (*indels*) nos genomas dos isolados geográficos, como evidenciado nas Figura 22 a Figura 38. Não foram realizadas análises mais detalhadas a nível de bases nesta região, no entanto, é possível confirmar a ausência de regiões com até 891 bp. A presença ou ausência das formas longas ou curtas deste *locus* foi utilizada como marcador φ (fi) de recombinação, caracterizando esta região como um *hotspot* de ganhos, perdas e trocas de segmentos genômicos entre genótipos de AgMNPV (CROIZIER & RIBEIRO, 1992).

5.4 A VARIABILIDADE GENÉTICA EM AgMNPV

A alta taxa de ocorrência de erros de sequenciamento em leituras de 454 gira em torno de 18% (HUSE et al., 2007), e tais erros podem superestimar as estimativas de diversidade intrapopulacional, uma vez que esta se utilizaria das informações nucleotídicas do impreciso dado bruto do sequenciamento (KUNIN et al., 2010). Para contornar esse problema e fazer uso das informações de variabilidade nucleotídica obtidas para os diferentes genes de AgMNPV, as análises de diversidade genética tiveram um enfoque interpopulacional, a partir das sequências gênicas consenso obtidas gene a gene em cada população. Na estratégia de montagem utilizada, a sobreposição entre dezenas de leituras amenizam consideravelmente os efeitos negativos gerados pelos erros de sequenciamento, resultando em sequências consenso com alta precisão (GOLDBERG et al., 2006; HUSE et al., 2007). Ademais, a existência de uma sequência referência (AgMNPV-2D) como parâmetro comparativo, e a verificação manual dos possíveis erros de alinhamento refinam satisfatoriamente a qualidade dos

resultados finais, de forma que estes possam ser utilizados em análises de diversidade genética.

As análises de diversidade apontaram que quase metade dos genes de AgMNPV (74 dos 152) possuem baixa diversidade genética, estando enquadrados na subcategoria de diversidade A1. Nesta subcategoria estão 9 genes que não apresentaram qualquer tipo de substituições nucleotídicas: Ag38; Ag109; Ag141; bro-a; bro-g; ctl-2; etm; iap3 e odve18. Entre os genes de A1 que apresentaram substituições (entre sinônimas e não-sinônimas), o θ variou entre 0,00031 (38k (ORF94, 938 bp): uma substituição, sendo esta responsável pela mudança de um aminoácido) e 0,0024 (Ag121 (ORF121, 729 bp): seis substituições, sendo uma responsável pela mudança de um aminoácido). Nesta subcategoria de diversidade encontram-se 22 dos 36 genes que codificam proteínas estruturais (funções "capsídeo" e "vírion"), dos quais 13 são compartilhados por todos os baculovírus (genes core) (Figura 62). As proteínas codificadas por estes genes estão envolvidas na formação das partículas virais, estabelecendo associações físicas entre si e com o DNA viral (ROHRMANN, 1992). Em função da esperada interação espacial entre proteínas estruturais durante a morfogênese, modificações em aminoácidos por ventura geradas em uma destas proteínas poderiam prejudicar as superfícies de interação destas com as demais proteínas estruturais, situação que favoreceria a lenta coevolução destes elementos, sob seleção purificadora (DAUGHERTY & MALIK, 2012). Esse ponto poderia explicar a baixa diversidade observada em proteínas estruturais, muitas das quais (em especial as codificadas por genes core) estão em constante coevolução desde o estabelecimento das interações entre estas nos ancestrais de AgMNPV.

As proteínas BRO-A e BRO-G possuem arquiteturas de domínios únicas em AgMNPV: a primeira é a única que possui o domínio T5orf172, enquanto a segunda é composta unicamente pelo domínio DUF3627, o qual possui baixa similaridade com domínios homólogos associados ao domínio Bro-N em outras proteínas BRO (Figura 66).

Diferentemente das categorias funcionais "vírion" e "capsídeo", compostas por genes menos diversos, as outras quatro categorias não apresentaram tendências claras de agrupamento quanto a níveis de conservação, estando mais uniformemente distribuídas entre as 4 categorias de diversidade.

As categorias funcionais dos genes foram definidas tendo como base nas suas principais funções segundo a literatura (COHEN *et al.*, 2009; GARAVAGLIA *et al.*, 2012; MIELE *et al.*, 2011; ROHRMANN, 2011; VAN OERS & VLAK, 2007), logo, as prováveis funções secundárias desempenhadas por estes genes não foram consideradas, fator que poderia introduzir um viés nas análises com base em categorias funcionais.

Outra observação derivada destas análises foi a baixa variabilidade da maioria das novas ORFs de AgMNPV (Figura 60 e Tabela 9). Entre as 29 ORFs, 26 possuem sobreposições com ORFs validadas ou regiões homólogas, elementos que naturalmente sofrem pressões evolutivas que restringem a fixação de mutações deletérias. Devido a estas sobreposições, é difícil inferir se as supostas ORFs apresentam baixa variabilidade por de fato apresentarem função codificante, ou simplesmente em razão de compartilharem bases com regiões conservadas. A origem *de novo* de proteínas com variadas funções e estruturas a partir de ORFs sobrepostas é um fenômeno comum entre vírus (RANCUREL et al., 2009). As evidências aqui apresentadas podem abrir precedentes para a identificação de proteínas codificadas dentro de contextos ainda mais complexos de organização gênica, como a possível existência de ORFs funcionais aninhadas em genes de baculovírus.

Quanto aos 17 genes com maior diversidade genética (incluídos nas subcategorias de B a D), 10 estão alocados especialmente numa pequena região genômica entre as hr4 e hr1, ou próximos a estas regiões homólogas. Como discutido anteriormente, este *locus* é um *hotspot* de mutações e recombinações entre genomas (CROIZIER & RIBEIRO, 1992; MARUNIAK *et al.*, 1999), e nela estão localizados diversos genes cujos produtos atuam como transativadores ou intensificadores transcricionais, como Ag4; Ag5 (LO et al., 2002); *ie*-0; *ie-1* (KOOL et al., 1994); *ie-2* (PRIKHOD'KO et al., 1999); *e pe-38* (KRAPPA *et al.*, 1995). Nesta região também são encontrados quatro genes *bro* (*a*, *b*, *c*, *e*), muitos dos quais codificam proteínas com provável capacidade de ligação a DNA e ação reguladora da transcrição de genes virais e de hospedeiros (em decorrência do domínio Bro-N) (ZEMSKOV *et al.*, 2000). A alta diversidade estrutural e nucleotídica desta região; seus recorrentes eventos de recombinação; assim como a grande concentração de elementos reguladores da transcrição indicam que suas diferentes configurações exercem papel determinante sobre a infectividade viral, uma vez que seus elementos genéticos (genes e *hrs*) coordenam mecanismos centrais da interação vírus-hospedeiro (BLISSARD & ROHRMANN, 1990).

6 CONCLUSÕES

- (i) A alta cobertura e a grande extensão média das leituras (entre 250 e 500 bp) em sequenciamentos em larga escala 454 Roche[®] garantem a montagem de genomas de baculovírus com grandes chances de recuperação da ligação genética entre polimorfismos de haplótipos curtos. Porém, a reconstrução global de genótipos individuas mostra-se impraticável, principalmente em sistemas que evoluem por recombinação homóloga;
- (ii) Erros de sequenciamento inerentes à tecnologia 454 Roche[®], em especial os relativos a inserções de Adeninas em regiões homopoliméricas, dificultam o processo de anotação genômica e as estimativas de diversidade genética intrapopulacional;
- (iii) Ao longo da histórica evolutiva de AgMNPV, eventos de inversão ou translocação de regiões genômicas não foram observados, o que sugere que a sintenia possa ser importante para modular a expressão gênica e definir a infectividade viral frente a diferentes hospedeiros;
- (iv) Foram identificados até 27 ORFs de supostos novos genes localizados em regiões intergênicas e sobrepostas a ORFs pré-anotadas no isolado 2D. A presença destas não são evidências suficientes para assegurar a sua funcionalidade como regiões codificadoras de proteínas. Por outro lado, a conservação da maioria destas ORFs nos 17 isolados geográficos sugere a existência de alguma pressão evolutiva que mantém as regiões intergênicas pouco variáveis;
- (v) A descoberta de uma possível ORF aninhada a um gene de baculovírus (*fgf*) abre precedentes para a possível existência de novos níveis de organização gênica ainda não descritos para baculovírus;
- (vi) Alguns genes de AgMNPV sofrem severas mudanças estruturais em função de substituições, deleções e inserções de um ou poucos nucleotídeos, eventos que causam mudanças de fase e levam a sobreposições, fissões ou fusões gênicas. Deleções e inserções de nucleotídeos em múltiplos de três também foram

observadas em alguns genes, tendo por consequência alterações estruturais ou truncamento de domínios protéicos;

- (vii) A ausência de bro-a em algumas populações dão indícios sobre como tais genes (presentes em múltiplas cópias) podem ter sido adquiridos e perdidos nos genomas de AgMNPV. Ademais, a presença de bro-a em alguns isolados, assim como a diversidade estrutural observada em bro-b e bro-c, reforçam a hipótese de participação das proteínas BRO como fatores determinantes da infectividade viral frente a diferentes hospedeiros;
- (viii) Um suposto novo gene (Ag25b) que codifica uma proteína com domínio RING Finger está presente em alguns genótipos de AgMNPV, mas pode ter sido perdido em alguns isolados geográficos deste vírus. O produto deste gene provavelmente atua na via de ubiquitinação (como uma ligase proteína-ubiquitina (E3)) em associação com o produto do gene adjacente, *ubi* (ORF26), que codifica uma ubiquitina viral;
- (ix) A maioria dos genes de baculovírus é pouco variável. Entre os genes mais conservados estão os genes *core*, em especial aqueles cujos produtos estão associados à formação de partículas virais (genes estruturais, de categorias funcionais "capsídeo" e "vírion").

REFERÊNCIAS*

ABBOTT, J. C.; AANENSEN, D. M.; BENTLEY, S. D. WebACT: an online genome comparison suite. **Methods Mol. Biol.**, v. 395, p. 57-74, 2007.

ADELMAN, J. P.; BOND, C. T.; DOUGLASS, J.; HERBERT, E. Two mammalian genes transcribed from opposite strands of the same DNA locus. **Science**, v. 235, n. 4795, p. 1514-1517, 1987.

ADHYA, S.; GOTTESMAN, M. Promoter occlusion: transcription through a promoter may inhibit its activity. **Cell**, v. 29, n. 3, p. 939-944, 1982.

AFONSO, C. L.; TULMAN, E. R.; LU, Z.; OMA, E.; KUTISH, G. F.; ROCK, D. L. The genome of Melanoplus sanguinipes entomopoxvirus. **J. Virol.**, v. 73, n. 1, p. 533-552, 1999.

ALLEN, G. E.; KNELL, J. D. A nuclear polyhedrosis virus of Anticarsia gemmatalis: I, ultrastructure, replication, and pathogenicity. **Fl. Entomol.**, n. 60, p. 233-240, 1977.

APIC, G.; GOUGH, J.; TEICHMANN, S. A. An insight into domain combinations. **Bioinformatics**, v. 17 Suppl 1, p. 83-89, 2001.

ARENDS, H. M.; JEHLE, J. A. Homologous recombination between the inverted terminal repeats of defective transposon TCp3.2 causes an inversion in the genome of Cydia pomonella granulovirus. **J. Gen. Virol.**, v. 83, n. Pt 7, p. 1573-1578, 2002.

ASSER-KAISER, S.; FRITSCH, E.; UNDORF-SPAHN, K.; KIENZLE, J.; EBERLE, K. E.; GUND, N. A.; REINEKE, A.; ZEBITZ, C. P.; HECKEL, D. G.; HUBER, J.; JEHLE, J. A. Rapid emergence of baculovirus resistance in codling moth due to dominant, sex-linked inheritance. **Science**, v. 317, n. 5846, p. 1916-1918, 2007.

AYRES, M. D.; HOWARD, S. C.; KUZIO, J.; LOPEZ-FERBER, M.; POSSEE, R. D. The complete DNA sequence of Autographa californica nuclear polyhedrosis virus. **Virology**, v. 202, n. 2, p. 586-605, 1994.

BECKER, D.; KNEBEL-MORSDORF, D. Sequence and temporal appearance of the early transcribed baculovirus gene HE65. J. Virol., v. 67, n. 10, p. 5867-5872, 1993.

BEERENWINKEL, N.; GUNTHARD, H. F.; ROTH, V.; METZNER, K. J. Challenges and opportunities in estimating viral genetic diversity from next-generation sequencing data. **Front. Microbiol.**, v. 3, p. 329, 2012.

BEERENWINKEL, N.; ZAGORDI, O. Ultra-deep sequencing for the analysis of viral populations. **Curr. Opin. Virol.**, v. 1, n. 5, p. 413-418, 2011.

BELOV, A. A.; MOHAMMADI, M. Molecular mechanisms of fibroblast growth factor signaling in physiology and pathology. **Cold Spring Harb. Perspect. Biol.**, v. 5, n. 6, p. 111, 2013.

^{*} De acordo com: ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. NBR 6023: Informação e documentação: referências: elaboração. Rio de janeiro, 2002.

BIDESHI, D. K.; RENAULT, S.; STASIAK, K.; FEDERICI, B. A.; BIGOT, Y. 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.**, v. 84, n. Pt 9, p. 2531-2544, 2003.

BILIMORIA, S. L. The biology of nuclear polyhedrosis viruses. In: KURSTAK, E. (Ed.). Viruses of invertebrates. New York: Marcel Dekker, INC, 1991. p.1-72.

BLANCA, J.; CAÑIZARES, J. sff_extract — Bioinformatics at COMAV 0.1 documentation. 2012. Disponível em: < http://bioinf.comav.upv.es/sff_extract/index.html >. Acesso em: 19 set. 2012.

BLISSARD, G. W.; KOGAN, P. H.; WEI, R.; ROHRMANN, G. F. A synthetic early promoter from a baculovirus: roles of the TATA box and conserved start site CAGT sequence in basal levels of transcription. **Virology**, v. 190, n. 2, p. 783-793, 1992.

BLISSARD, G. W.; ROHRMANN, G. F. Baculovirus diversity and molecular biology. Annu. Rev. Entomol., v. 35, p. 127-155, 1990.

BOND, C. T.; HAYFLICK, J. S.; SEEBURG, P. H.; ADELMAN, J. P. The rat gonadotropinreleasing hormone: SH locus: structure and hypothalamic expression. **Mol. Endocrinol.**, v. 3, n. 8, p. 1257-1262, 1989.

BONNING, B. C. Baculoviruses: biology, biochemistry and molecular biology. In: GILBERT, L. I.;IATROU, K. e GILL, S. S. (Ed.). **Comprehensive molecular insect science**. Oxford: Elsevier, 2005. p.233-270.

BORST, P.; GREAVES, D. R. Programmed gene rearrangements altering gene expression. **Science**, v. 235, n. 4789, p. 658-667, 1987.

BROWN, S. E.; MARUNIAK, J. E.; KNUDSON, D. L. Baculovirus (MNPV) genomic variants: characterization of Spodoptera exempta MNPV DNAs and comparison with other Autographa californica MNPV DNAs. **J. Gen. Virol.**, v. 66 (Pt 11), p. 2431-2441, 1985.

CARSON, D. D.; SUMMERS, M. D.; GUARINO, L. A. Molecular analysis of a baculovirus regulatory gene. **Virology**, v. 182, n. 1, p. 279-286, 1991.

CARVER, T. J.; RUTHERFORD, K. M.; BERRIMAN, M.; RAJANDREAM, M. A.; BARRELL, B. G.; PARKHILL, J. ACT: the Artemis Comparison Tool. **Bioinformatics**, v. 21, n. 16, p. 3422-3423, 2005.

CASTRO, M. E.; SOUZA, M. L.; ARAUJO, S.; BILIMORIA, S. L. Replication of Anticarsia gemmatalis nuclear polyhedrosis virus in four lepidopteran cell lines. **J. Invertebr. Pathol.**, v. 69, n. 1, p. 40-45, 1997.

CHEN, C. N.; MALONE, T.; BECKENDORF, S. K.; DAVIS, R. L. At least two genes reside within a large intron of the dunce gene of Drosophila. **Nature**, v. 329, n. 6141, p. 721-724, 1987.

CHEVREUX, B.; WETTER, T.; SUHAI, S. Genome Sequence Assembly Using Trace Signals and Additional Sequence Information. German Conference on Bioinformatics, 1999. Hannover, Germany. p.45-56.

CHIRICO, N.; VIANELLI, A.; BELSHAW, R. Why genes overlap in viruses. **Proc. Biol. Sci.**, v. 277, n. 1701, p. 3809-3817, 2010.

CHISHOLM, G. E.; HENNER, D. J. Multiple early transcripts and splicing of the Autographa californica nuclear polyhedrosis virus IE-1 gene. J. Virol., v. 62, n. 9, p. 3193-3200, 1988.

CHOI, J.; GUARINO, L. A. The baculovirus transactivator IE1 binds to viral enhancer elements in the absence of insect cell factors. **J. Virol.**, v. 69, n. 7, p. 4548-4551, 1995.

CLEM, R. J. Baculoviruses and apoptosis: the good, the bad, and the ugly. Cell Death Differ., v. 8, n. 2, p. 137-143, 2001.

CLEM, R. J.; MILLER, L. K. Control of programmed cell death by the baculovirus genes p35 and iap. **Mol. Cell. Biol.**, v. 14, n. 8, p. 5212-5222, 1994.

COCHRAN, M. A.; BROWN, S. E.; KNUDSON, D. L. Organization and expression of the baculovirus genome. In: GRANADOS, R. R. e FEDERICI, B. A. (Ed.). **The biology of baculovirus**. Florida: CRC Press, v.I, 1986. p.239-258.

COCHRAN, M. A.; FAULKNER, P. Location of Homologous DNA Sequences Interspersed at Five Regions in the Baculovirus AcMNPV Genome. **J. Virol.**, v. 45, n. 3, p. 961-970, 1983.

COHEN, D.; MAREK, M.; DAVIES, B.; VLAK, J. M.; VAN OERS, M. Encyclopedia of Autographa californica Nucleopolyhedrovirus Genes. **Virol. Sin.**, v. 24, n. 5, p. 359-414, 2009.

COOPER, D.; CORY, J. S.; MYERS, J. H. Hierarchical spatial structure of genetically variable nucleopolyhedroviruses infecting cyclic populations of western tent caterpillars. **Mol. Ecol.**, v. 12, n. 4, p. 881-890, 2003.

CORRADI, N.; GANGAEVA, A.; KEELING, P. J. Comparative profiling of overlapping transcription in the compacted genomes of microsporidia Antonospora locustae and Encephalitozoon cuniculi. **Genomics**, v. 91, n. 4, p. 388-393, 2008.

CROIZIER, G.; RIBEIRO, H. C. T. Recombination as a possible major cause of genetic heterogeneity in Anticarsia gemmatalis nuclear polyhedrosis virus wild populations. **Virus Res.**, v. 26, p. 183-196, 1992.

DANFORD, T.; DOWELL, R.; AGARWALA, S.; GRISAFI, P.; FINK, G.; GIFFORD, D. Discovering regulatory overlapping RNA transcripts. **J. Comput. Biol.**, v. 18, n. 3, p. 295-303, 2011.

DAUGHERTY, M. D.; MALIK, H. S. Rules of engagement: molecular insights from hostvirus arms races. **Annu. Rev. Genet.**, v. 46, p. 677-700, 2012. DINGER, M. E.; GASCOIGNE, D. K.; MATTICK, J. S. The evolution of RNAs with multiple functions. **Biochimie**, v. 93, n. 11, p. 2013-2018, 2011.

DUFFY, S.; SHACKELTON, L. A.; HOLMES, E. C. Rates of evolutionary change in viruses: patterns and determinants. **Nat. Rev. Genet.**, v. 9, n. 4, p. 267-276, 2008.

ENTWISTLE, P. F.; ADAMS, P. H. W.; EVANS, H. F.; RIVERS, C. F. Epizootiology of a nuclear polyhedrosis virus (Baculoviridae) in European spruce sawfly (Gilpinia hercyniae): spread of disease from small epicentres in comparison with spread of baculovirus diseases in other hosts. J. Appl. Ecol., v. 20, p. 473–487, 1983.

ERIKSSON, N.; PACHTER, L.; MITSUYA, Y.; RHEE, S. Y.; WANG, C.; GHARIZADEH, B.; RONAGHI, M.; SHAFER, R. W.; BEERENWINKEL, N. Viral population estimation using pyrosequencing. **PLoS Comput. Biol.**, v. 4, n. 4, p. e1000074, 2008.

FLIPSEN, J. T.; MANS, R. M.; KLEEFSMAN, A. W.; KNEBEL-MORSDORF, D.; VLAK, J. M. Deletion of the baculovirus ecdysteroid UDP-glucosyltransferase gene induces early degeneration of Malpighian tubules in infected insects. **J. Virol.**, v. 69, n. 7, p. 4529-4532, 1995.

FREEMONT, P. S. Ubiquitination: RING for destruction? Curr. Biol., v. 10, n. 2, p. R84-87, 2000.

FRIESEN, P. D.; MILLER, L. K. Temporal regulation of baculovirus RNA: overlapping early and late transcripts. **J. Virol.**, v. 54, n. 2, p. 392-400, 1985.

_____. The regulation of baculovirus gene expression. In: DERFLER, W. e BOEHM, P. (Ed.). **The molecular biology of baculoviruses**. Berlin: Springer-Verlag, 1986. p.31-50.

_____. Insect viruses. In: KNIPE, D. M.;HOWLEY, P. M.;GRIFFIN, D. E.;LAMB, R. A.;MARTIN, M. A.;ROIZMAN, B. e STRAUS, S. E. (Ed.). Field's virology. 4. Philadelphia: Lippincott Williams & Wilkins, 2001.

FUNK, C. J.; BRAUNAGEL, S. C.; ROHRMANN, G. F. Baculovirus structure. In: MILLER, L. K. (Ed.). **The baculoviruses**. New York: Springer, 1997. p.7-32.

FURIA, M.; D'AVINO, P. P.; CRISPI, S.; ARTIACO, D.; POLITO, L. C. Dense cluster of genes is located at the ecdysone-regulated 3C puff of Drosophila melanogaster. **J. Mol. Biol.**, v. 231, n. 2, p. 531-538, 1993.

FUXA, J. R. Ecology of insect nucleopolyhedroviruses. Agric. Ecosyst. Environ., v. 103, p. 27–43, 2004.

FUXA, J. R.; RICHTER, A. R. Classical biological control in an ephemeral crop habitat with Anticarsia gemmatalis nucleopolyhedrovirus. **BioControl**, v. 44, p. 403–419, 1999.

GARAVAGLIA, M. J.; MIELE, S. A.; ISERTE, J. A.; BELAICH, M. N.; GHIRINGHELLI, P. D. The ac53, ac78, ac101, and ac103 genes are newly discovered core genes in the family Baculoviridae. **J. Virol.**, v. 86, n. 22, p. 12069-12079, 2012.

GARCIA-MARUNIAK, A.; PAVAN, O. H.; MARUNIAK, J. E. A variable region of Anticarsia gemmatalis nuclear polyhedrosis virus contains tandemly repeated DNA sequences. **Virus Res.**, v. 41, n. 2, p. 123-132, 1996.

GELERNTER, W. D.; FEDERICI, B. A. Virus epizootics in Californian populations of Spodoptera exigua: dominance of a single viral genotype. **Biochem. System. Ecol.**, v. 18, p. 461–466, 1990.

GETTIG, R. R.; MCCARTHY, W. J. Genotypic variation among wild isolates of Heliothis spp nuclear polyhedrosis viruses from different geographical regions. **Virology**, v. 117, n. 1, p. 245-252, 1982.

GIBSON, C. W.; THOMSON, N. H.; ABRAMS, W. R.; KIRKHAM, J. Nested genes: biological implications and use of AFM for analysis. **Gene**, v. 350, n. 1, p. 15-23, 2005.

GOLDBERG, S. M.; JOHNSON, J.; BUSAM, D.; FELDBLYUM, T.; FERRIERA, S.; FRIEDMAN, R.; HALPERN, A.; KHOURI, H.; KRAVITZ, S. A.; LAURO, F. M.; LI, K.; ROGERS, Y. H.; STRAUSBERG, R.; SUTTON, G.; TALLON, L.; THOMAS, T.; VENTER, E.; FRAZIER, M.; VENTER, J. C. A Sanger/pyrosequencing hybrid approach for the generation of high-quality draft assemblies of marine microbial genomes. **Proc. Natl. Acad. Sci. USA,** v. 103, n. 30, p. 11240-11245, 2006.

GORDON, D.; ABAJIAN, C.; GREEN, P. Consed: a graphical tool for sequence finishing. **Genome Res.**, v. 8, n. 3, p. 195-202, 1998.

GRANADOS, R. R.; LAWLER, K. A. In vivo pathway of Autographa californica baculovirus invasion and infection. **Virology**, v. 108, n. 2, p. 297-308, 1981.

GREENE, G. L.; LEPPLA, N. C.; DICKERSON, W. A. Velvetbean caterpillar: a rearing procedure and artificial medium. **J. Econ. Entomol.**, v. 69, n. 4, p. 487-488, 1976.

GRÖNER, A. Specificity and safety of baculoviruses. In: GRANADOS, R. R. e FEDERICI, B. A. (Ed.). **The biology of baculoviruses**. Boca Raton: CRC Press, v.1, 1986. p.177-202.

GUARINO, L. A. Identification of a viral gene encoding a ubiquitin-like protein. **Proc. Natl.** Acad. Sci. USA, v. 87, n. 1, p. 409-413, 1990.

GUARINO, L. A.; GONZALEZ, M. A.; SUMMERS, M. D. Complete Sequence and Enhancer Function of the Homologous DNA Regions of Autographa californica Nuclear Polyhedrosis Virus. J. Virol., v. 60, n. 1, p. 224-229, 1986.

GUARINO, L. A.; XU, B.; JIN, J.; DONG, W. A virus-encoded RNA polymerase purified from baculovirus-infected cells. J. Virol., v. 72, n. 10, p. 7985-7991, 1998.

HAJOS, J. P.; PIJNENBURG, J.; USMANY, M.; ZUIDEMA, D.; ZAVODSZKY, P.; VLAK, J. M. High frequency recombination between homologous baculoviruses in cell culture. **Arch. Virol.**, v. 145, n. 1, p. 159-164, 2000.

HALL, T. A. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. **Nucl. Acids. Symp. Ser.**, n. 41, p. 95-98, 1999.

HAPP, B.; LI, J.; DOERFLER, W. Proteins encoded in the 81.2- to 85.0-map-unit fragment of Autographa californica nuclear polyhedrosis virus DNA can be translated in vitro and in Spodoptera frugiperda cells. **J. Virol.**, v. 65, n. 1, p. 89-97, 1991.

HARRAP, K. A. The structure of nuclear polyhedrosis viruses. II. The virus particle. **Virology**, v. 50, n. 1, p. 124-132, 1972.

HARRISON, R. L. Structural divergence among genomes of closely related baculoviruses and its implications for baculovirus evolution. **J. Invertebr. Pathol.,** v. 101, n. 3, p. 181-186, 2009.

HARRISON, R. L.; BONNING, B. C. The nucleopolyhedroviruses of Rachiplusia ou and Anagrapha falcifera are isolates of the same virus. **J. Gen. Virol.**, v. 80 (Pt 10), p. 2793-2798, 1999.

HAYAKAWA, T.; ROHRMANN, G. F.; HASHIMOTO, Y. Patterns of genome organization and content in lepidopteran baculoviruses. **Virology**, v. 278, n. 1, p. 1-12, 2000.

HEDRICK, P. W. Genetics of populations. Sudbury, Boston, Toronto, London, Singapore: Jones and Bartlett Publishers, 2005.

HENIKOFF, S.; KEENE, M. A.; FECHTEL, K.; FRISTROM, J. W. Gene within a gene: nested Drosophila genes encode unrelated proteins on opposite DNA strands. **Cell**, v. 44, n. 1, p. 33-42, 1986.

HERNIOU, E. A.; LUQUE, T.; CHEN, X.; VLAK, J. M.; WINSTANLEY, D.; CORY, J. S.; O'REILLY, D. R. Use of whole genome sequence data to infer baculovirus phylogeny. J. Virol., v. 75, n. 17, p. 8117-8126, 2001.

HERNIOU, E. A.; OLSZEWSKI, J. A.; CORY, J. S.; O'REILLY, D. R. The genome sequence and evolution of baculoviruses. **Annu. Rev. Entomol.**, v. 48, p. 211-234, 2003.

HODGSON, D. J.; HITCHMAN, R. B.; VANBERGEN, A. J.; HAILS, R. S.; POSSEE, R. D.; CORY, J. S. Host ecology determines the relative fitness of virus genotypes in mixed-genotype nucleopolyhedrovirus infections. **J. Evol. Biol.**, v. 17, n. 5, p. 1018-1025, 2004.

HODGSON, D. J.; VANBERGEN, A. J.; WATT, A. D.; HAILS, R. S.; CORY, J. S. Phenotypic variation between naturally co-existing genotypes of a Lepidopteran baculovirus. **Evol. Ecol. Res.**, v. 3, p. 687–701, 2001.

HOSTETTER, D. L.; PUTTLER, B. A new broad spectrum nuclear polyhedrosis virus isolated from the celery looper, Anagrapha falcifera (Kirby), (Lepidoptera: Noctuidae). **Environ. Entomol.**, v. 20, p. 1480-1488, 1991.

HUGHES, A. L. Evolution of inhibitors of apoptosis in baculoviruses and their insect hosts. **Infect. Genet. Evol.,** v. 2, n. 1, p. 3-10, 2002a.

_____. Origin and evolution of viral interleukin-10 and other DNA virus genes with vertebrate homologues. **J. Mol. Evol.,** v. 54, n. 1, p. 90-101, 2002b.

HUGHES, A. L.; FRIEDMAN, R. Genome-wide survey for genes horizontally transferred from cellular organisms to baculoviruses. **Mol. Biol. Evol.**, v. 20, n. 6, p. 979-987, 2003.

HUGHES, T. A. Regulation of gene expression by alternative untranslated regions. **Trends Genet.**, v. 22, n. 3, p. 119-122, 2006.

HUH, N. E.; WEAVER, R. F. Identifying the RNA polymerases that synthesize specific transcripts of the Autographa californica nuclear polyhedrosis virus. **J. Gen. Virol.**, v. 71 (Pt 1), p. 195-201, 1990.

HURST, L. D.; WILLIAMS, E. J.; PAL, C. Natural selection promotes the conservation of linkage of co-expressed genes. **Trends Genet.**, v. 18, n. 12, p. 604-606, 2002.

HUSE, S. M.; HUBER, J. A.; MORRISON, H. G.; SOGIN, M. L.; WELCH, D. M. Accuracy and quality of massively parallel DNA pyrosequencing. **Genome Biol.**, v. 8, n. 7, p. R143, 2007.

HYINK, O.; DELLOW, R. A.; OLSEN, M. J.; CARADOC-DAVIES, K. M.; DRAKE, K.; HERNIOU, E. A.; CORY, J. S.; O'REILLY, D. R.; WARD, V. K. Whole genome analysis of the Epiphyas postvittana nucleopolyhedrovirus. **J. Gen. Virol.**, v. 83, n. Pt 4, p. 957-971, 2002.

IJKEL, W. F.; VAN STRIEN, E. A.; HELDENS, J. G.; BROER, R.; ZUIDEMA, D.; GOLDBACH, R. W.; VLAK, J. M. Sequence and organization of the Spodoptera exigua multicapsid nucleopolyhedrovirus genome. **J. Gen. Virol.**, v. 80 (Pt 12), p. 3289-3304, 1999.

IMAI, N.; MATSUDA, N.; TANAKA, K.; NAKANO, A.; MATSUMOTO, S.; KANG, W. Ubiquitin ligase activities of Bombyx mori nucleopolyhedrovirus RING finger proteins. J. Virol., v. 77, n. 2, p. 923-930, 2003.

IYER, L. M.; KOONIN, E. V.; ARAVIND, L. Extensive domain shuffling in transcription regulators of DNA viruses and implications for the origin of fungal APSES transcription factors. **Genome Biol.**, v. 3, n. 3, p. RESEARCH0012, 2002.

JAKUBOWSKA, A. K.; PETERS, S. A.; ZIEMNICKA, J.; VLAK, J. M.; VAN OERS, M. M. Genome sequence of an enhancin gene-rich nucleopolyhedrovirus (NPV) from Agrotis segetum: collinearity with Spodoptera exigua multiple NPV. **J. Gen. Virol.**, v. 87, n. Pt 3, p. 537-551, 2006.

JAKUBOWSKA, A. K.; VOGEL, H.; HERRERO, S. Increase in Gut Microbiota after Immune Suppression in Baculovirus-infected Larvae. **PLoS Pathog.**, v. 9, n. 5, p. e1003379, 2013.

JAQUES, R. P. The persistence of a nuclear polyhedrosis virus in the habitat of the host insect, Trichoplusia ni. I. Polyhedra deposited on foliage. **Can. Entomol.,** v. 9, p. 785–794, 1967.

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. On the

classification and nomenclature of baculoviruses: a proposal for revision. Arch. Virol., v. 151, n. 7, p. 1257-1266, 2006.

JIN, J.; GUARINO, L. A. 3'-end formation of baculovirus late RNAs. J. Virol., v. 74, n. 19, p. 8930-8937, 2000.

JOAZEIRO, C. A.; WEISSMAN, A. M. RING finger proteins: mediators of ubiquitin ligase activity. **Cell**, v. 102, n. 5, p. 549-552, 2000.

JOHNSON, D. W.; MARUNIAK, J. E. Physical map of Anticarsia-gemmatalis nuclear polyhedrosis-virus (AgMNPV-2) DNA. J. Gen. Virol., n. 70, p. 1877-1883, 1989.

JOJIC, V.; HERTZ, T.; JOJIC, N. Population sequencing using short reads: HIV as a case study. **Pac. Symp. Biocomput.**, p. 114-125, 2008.

JONES, K. A. Bioassays of entomopathogenic viruses. In: NAVON, A. e ASCHER, K. R. S. (Ed.). **Bioassays of entomopathogenic microbes and nematodes**. Orford: CABI, 2000. cap. 3, p.324.

KAMITA, S. G.; MAEDA, S.; HAMMOCK, B. D. High-frequency homologous recombination between baculoviruses involves DNA replication. J. Virol., v. 77, n. 24, p. 13053-13061, 2003.

KANG, W.; SUZUKI, M.; ZEMSKOV, E.; OKANO, K.; MAEDA, S. Characterization of baculovirus repeated open reading frames (bro) in Bombyx mori nucleopolyhedrovirus. J. Virol., v. 73, n. 12, p. 10339-10345, 1999.

KATSUMA, S.; KANG, W.; SHIN-I, T.; OHISHI, K.; KADOTA, K.; KOHARA, Y.; SHIMADA, T. Mass identification of transcriptional units expressed from the Bombyx mori nucleopolyhedrovirus genome. **J. Gen. Virol.**, v. 92, n. Pt 1, p. 200-203, 2011.

KELLER, A.; NESVIZHSKII, A. I.; KOLKER, E.; AEBERSOLD, R. Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. **Anal Chem.**, v. 74, n. 20, p. 5383-5392, 2002.

KIMELMAN, D.; KIRSCHNER, M. W. An antisense mRNA directs the covalent modification of the transcript encoding fibroblast growth factor in Xenopus oocytes. **Cell**, v. 59, n. 4, p. 687-696, 1989.

KISLEV, N.; EDELMAN, M. DNA restriction-pattern differences from geographic isolates of Spodoptera littoralis nuclear polyhedrosis virus. **Virology**, v. 119, n. 1, p. 219-222, 1982.

KONDO, A.; MAEDA, S. Host range expansion by recombination of the baculoviruses Bombyx mori nuclear polyhedrosis virus and Autographa californica nuclear polyhedrosis virus. **J. Virol.**, v. 65, n. 7, p. 3625-3632, 1991.

KOOL, M.; AHRENS, C. H.; GOLDBACH, R. W.; ROHRMANN, G. F.; VLAK, J. M. Identification of genes involved in DNA replication of the Autographa californica baculovirus. **Proc. Natl. Acad. Sci. USA**, v. 91, n. 23, p. 11212-11216, 1994.

KOOL, M.; VLAK, J. M. The structural and functional organization of the Autographa californica nuclear polyhedrosis virus genome. **Arch. Virol.**, v. 130, n. 1-2, p. 1-16, 1993.

KRAKAUER, D. C. Stability and evolution of overlapping genes. **Evolution**, v. 54, n. 3, p. 731-739, 2000.

KRAKAUER, D. C.; ZANOTTO, P. M. A. Viral individuality and limitations of the life concept. In: RASMUSSEN, S.;BEDAU, M. A.;CHEN, L.;DEAMER, D.;KRAKAUER, D. C.;PACKARD, N. e STADLER, D. P. (Ed.). **Protocells: bridging nonliving and living matter**. Hong Kong: MIT Press 2009. p.513-536.

KRAPPA, R.; KNEBEL-MORSDORF, D. Identification of the very early transcribed baculovirus gene PE-38. J. Virol., v. 65, n. 2, p. 805-812, 1991.

KRAPPA, R.; RONCARATI, R.; KNEBEL-MORSDORF, D. Expression of PE38 and IE2, viral members of the C3HC4 finger family, during baculovirus infection: PE38 and IE2 localize to distinct nuclear regions. **J. Virol.**, v. 69, n. 9, p. 5287-5293, 1995.

KUMMERFELD, S. K.; TEICHMANN, S. A. Relative rates of gene fusion and fission in multi-domain proteins. **Trends Genet.**, v. 21, n. 1, p. 25-30, 2005.

KUNIMI, Y.; FUXA, J. R.; HAMMOCK, B. D. Comparison of wild type and genetically engineered nuclear polyhedrosis

viruses of Autographa californica for mortality, virus replication and polyhedra production in Trichoplusia ni larvae. **Entomol. Exp. Appl.**, v. 81, p. 129–135, 1996.

KUNIN, V.; ENGELBREKTSON, A.; OCHMAN, H.; HUGENHOLTZ, P. Wrinkles in the rare biosphere: pyrosequencing errors can lead to artificial inflation of diversity estimates. **Environ. Microbiol.**, v. 12, n. 1, p. 118-123, 2010.

KUZIO, J.; PEARSON, M. N.; HARWOOD, S. H.; FUNK, C. J.; EVANS, J. T.; SLAVICEK, J. M.; ROHRMANN, G. F. Sequence and analysis of the genome of a baculovirus pathogenic for Lymantria dispar. **Virology**, v. 253, n. 1, p. 17-34, 1999.

LANDSCHULZ, W. H.; JOHNSON, P. F.; MCKNIGHT, S. L. The leucine zipper: a hypothetical structure common to a new class of DNA binding proteins. **Science**, v. 240, n. 4860, p. 1759-1764, 1988.

LANGE, M.; JEHLE, J. A. The genome of the Cryptophlebia leucotreta granulovirus. **Virology**, v. 317, n. 2, p. 220-236, 2003.

LAUZON, H. A.; JAMIESON, P. B.; KRELL, P. J.; ARIF, B. M. Gene organization and sequencing of the Choristoneura fumiferana defective nucleopolyhedrovirus genome. J. Gen. Virol., v. 86, n. Pt 4, p. 945-961, 2005.

LEE, H. H.; MILLER, L. K. Isolation of genotypic variants of Autographa californica nuclear polyhedrosis virus. **J. Virol.**, v. 27, n. 3, p. 754-767, 1978.

_____. Isolation, Complementation, and Initial Characterization of Temperature-Sensitive Mutants of the Baculovirus Autographa californica Nuclear Polyhedrosis Virus. J. Virol., v. 31, n. 1, p. 240-252, 1979.

LEE, Y.; FUXA, J. R. Transport of Wild-Type and Recombinant Nucleopolyhedroviruses by Scavenging and Predatory Arthropods. **Microb. Ecol.**, v. 39, n. 4, p. 301-313, 2000.

LEVERSON, J. D.; JOAZEIRO, C. A.; PAGE, A. M.; HUANG, H.; HIETER, P.; HUNTER, T. The APC11 RING-H2 finger mediates E2-dependent ubiquitination. **Mol. Biol. Cell.**, v. 11, n. 7, p. 2315-2325, 2000.

LEVIN, D. B.; WHITTOME, B. Codon usage in nucleopolyhedroviruses. J. Gen. Virol., v. 81, n. Pt 9, p. 2313-2325, 2000.

LEVINSON, B.; KENWRICK, S.; LAKICH, D.; HAMMONDS, G., JR.; GITSCHIER, J. A transcribed gene in an intron of the human factor VIII gene. **Genomics**, v. 7, n. 1, p. 1-11, 1990.

LEVINSON, G.; GUTMAN, G. A. Slipped-strand mispairing: a major mechanism for DNA sequence evolution. **Mol. Biol. Evol.**, v. 4, n. 3, p. 203-221, 1987.

LI, L.; DONLY, C.; LI, Q.; WILLIS, L. G.; KEDDIE, B. A.; ERLANDSON, M. A.; THEILMANN, D. A. Identification and genomic analysis of a second species of nucleopolyhedrovirus isolated from Mamestra configurata. **Virology**, v. 297, n. 2, p. 226-244, 2002.

LIBRADO, P.; ROZAS, J. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. **Bioinformatics**, v. 25, n. 11, p. 1451-1452, 2009.

LILLO, F.; KRAKAUER, D. C. A statistical analysis of the three-fold evolution of genomic compression through frame overlaps in prokaryotes. **Biol. Direct,** v. 2, p. 22, 2007.

LO, H. R.; CHOU, C. C.; WU, T. Y.; YUEN, J. P.; CHAO, Y. C. Novel baculovirus DNA elements strongly stimulate activities of exogenous and endogenous promoters. **J. Biol. Chem.**, v. 277, n. 7, p. 5256-5264, 2002.

LOPEZ FERBER, M.; ARGAUD, O.; CROIZIER, L.; CROIZIER, G. Diversity, distribution, and mobility of bro gene sequences in Bombyx mori nucleopolyhedrovirus. **Virus Genes**, v. 22, n. 3, p. 247-254, 2001.

LORICK, K. L.; JENSEN, J. P.; FANG, S.; ONG, A. M.; HATAKEYAMA, S.; WEISSMAN, A. M. RING fingers mediate ubiquitin-conjugating enzyme (E2)-dependent ubiquitination. **Proc. Natl. Acad. Sci. USA**, v. 96, n. 20, p. 11364-11369, 1999.

LOVERING, R.; HANSON, I. M.; BORDEN, K. L.; MARTIN, S.; O'REILLY, N. J.; EVAN, G. I.; RAHMAN, D.; PAPPIN, D. J.; TROWSDALE, J.; FREEMONT, P. S. Identification and preliminary characterization of a protein motif related to the zinc finger. **Proc. Natl.** Acad. Sci. USA, v. 90, n. 6, p. 2112-2116, 1993.
LU, A.; MILLER, L. K. Identification of three late expression factor genes within the 33.8- to 43.4-map-unit region of Autographa californica nuclear polyhedrosis virus. **J. Virol.**, v. 68, n. 10, p. 6710-6718, 1994.

_____. Regulation of baculovirus late and very late gene expression. In: MILLER, L. K. (Ed.). **The baculoviruses**. New York: Plenum, 1997. p.193–216.

LUBBERT, H.; DOERFLER, W. Transcription of overlapping sets of RNAs from the genome of Autographa californica nuclear polyhedrosis virus: a novel method for mapping RNAs. J. Virol., v. 52, n. 1, p. 255-265, 1984.

LUQUE, T.; FINCH, R.; CROOK, N.; O'REILLY, D. R.; WINSTANLEY, D. The complete sequence of the Cydia pomonella granulovirus genome. **J. Gen. Virol.**, v. 82, n. Pt 10, p. 2531-2547, 2001.

MAEDA, S.; MUKOHARA, Y.; KONDO, A. Characteristically distinct isolates of the nuclear polyhedrosis virus from Spodoptera litura. **J. Gen. Virol.**, v. 71 (Pt 11), p. 2631-2639, 1990.

MAINPRIZE, T. H.; LEE, K.; MILLER, L. K. Variation in the temporal expression of overlapping baculovirus transcripts. **Virus Res.**, v. 6, n. 1, p. 85-99, 1986.

MARCHLER-BAUER, A.; LU, S.; ANDERSON, J. B.; CHITSAZ, F.; DERBYSHIRE, M. K.; DEWEESE-SCOTT, C.; FONG, J. H.; GEER, L. Y.; GEER, R. C.; GONZALES, N. R.; GWADZ, M.; HURWITZ, D. I.; JACKSON, J. D.; KE, Z.; LANCZYCKI, C. J.; LU, F.; MARCHLER, G. H.; MULLOKANDOV, M.; OMELCHENKO, M. V.; ROBERTSON, C. L.; SONG, J. S.; THANKI, N.; YAMASHITA, R. A.; ZHANG, D.; ZHANG, N.; ZHENG, C.; BRYANT, S. H. CDD: a Conserved Domain Database for the functional annotation of proteins. **Nucleic Acids Res.**, v. 39, n. Database issue, p. D225-229, 2011.

MARTIGNONI, M. E.; IWAI, P. J. A. **A catalog of viral diseases of insects, mites, and ticks**. 4th. Portland: General Technical Report PNW-40, 1986. 51 Disponível em: < http://www.fs.fed.us/pnw/publications/pnw_gtr195/pnw_gtr195a.pdf >.

MARTIN, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. **EMBnet.journal**, v. 17, n. 1, p. 10-12, 2011.

MARUNIAK, J. E. Baculovirus structural proteins and protein synthesis. In: GRANADOS, R. R. e FEDERICI, B. A. (Ed.). **The biology of baculovirus**. Florida: CRC Press, v.I, 1986. p.129-146.

_____. Molecular biology of Anticarsia gemmatalis baculovirus. **Mem. Inst. O. Cruz**, n. 84, p. 107-111, 1989.

MARUNIAK, J. E.; GARCIA-MARUNIAK, A.; SOUZA, M. L.; ZANOTTO, P. M.; MOSCARDI, F. Physical maps and virulence of Anticarsia gemmatalis nucleopolyhedrovirus genomic variants. **Arch. Virol.**, v. 144, n. 10, p. 1991-2006, 1999.

MATTICK, J. S. The double life of RNA. Biochimie, v. 93, n. 11, p. viii-ix, 2011.

MCINTOSH, A. H.; RICE, W. C.; IGNOFFO, C. M. Genotypic variants in wild-type populations of baculovirus. In: MARAMOROSCH, K. (Ed.). **Biotechnology in invertebrate pathology and cell culture**. New York: Academic Press, 1987. p.305-325.

MERRYWEATHER-CLARKE, A. T.; HIRST, M.; POSSEE, R. D. Recombination between genetically modified and unmodified Autographa californica nuclear polyhedrosis virus in Trichoplusia ni larvae. **Acta Virol.**, v. 38, n. 6, p. 311-315, 1994.

MIELE, S. A.; GARAVAGLIA, M. J.; BELAICH, M. N.; GHIRINGHELLI, P. D. Baculovirus: molecular insights on their diversity and conservation. Int. J. Evol. Biol., v. 2011, 2011.

MOSCARDI, F. Use of viruses for pest control in Brazil: the case of the nuclear polyhedrosis virus of the soybean caterpillar, anticarsia gemmatalis. **Mem. Inst. O. Cruz,** v. 84, p. 51-56, 1989.

_____. Assessment of the application of baculoviruses for control of Lepidoptera. Annu. Rev. Entomol., v. 44, p. 257-289, 1999.

MOSCARDI, F.; SOSA-GOMEZ, D. R. Use of viruses against soybean caterpillars in Brazil. In: COPPING, L. G.;GREEN, M. e REEDS, R. (Ed.). **Pest management in soybean**. London: Elsevier, 1992. p.98–109.

MOSCARDI, F.; SOUZA, M. L.; CASTRO, M. E.; MOSCARDI, M. L.; SZEWCZYK, B. Baculovirus pesticides: Present state and future perspectives. In: AHMAD, I.;AHMAD, F. e PICHTEL, J. (Ed.). Microbes and microbial technology. New York: Springer, 2011. p.415-445.

MUNOZ, D.; MURILLO, R.; KRELL, P. J.; VLAK, J. M.; CABALLERO, P. Four genotypic variants of a Spodoptera exigua Nucleopolyhedrovirus (Se-SP2) are distinguishable by a hypervariable genomic region. **Virus Res.**, v. 59, n. 1, p. 61-74, 1999.

MUNOZ, D.; VLAK, J. M.; CABALLERO, P. In vivo recombination between two strains of the genus Nucleopolyhedrovirus in its natural host, Spodoptera exigua. **Appl. Environ. Microbiol.**, v. 63, n. 8, p. 3025-3031, 1997.

MURPHY, P. R.; KNEE, R. S. Identification and characterization of an antisense RNA transcript (gfg) from the human basic fibroblast growth factor gene. **Mol. Endocrinol.,** v. 8, n. 7, p. 852-859, 1994.

NELSON, C. E.; HERSH, B. M.; CARROLL, S. B. The regulatory content of intergenic DNA shapes genome architecture. **Genome Biol.**, v. 5, n. 4, p. R25, 2004.

NORMARK, S.; BERGSTROM, S.; EDLUND, T.; GRUNDSTROM, T.; JAURIN, B.; LINDBERG, F. P.; OLSSON, O. Overlapping genes. **Annu. Rev. Genet.,** v. 17, p. 499-525, 1983.

O'REILLY, D. R. Auxiliary genes of baculoviruses. In: MILLER, L. K. (Ed.). The Baculoviruses. New York: Springer, 1997. cap. 3, p.267-300.

O'REILLY, D. R.; MILLER, L. K.; LUCKOW, V. A. **Baculovirus expression vectors: A laboratory manual**. New York: Oxford University Press, 1993. 368

OELLIG, C.; HAPP, B.; MULLER, T.; DOERFLER, W. Overlapping sets of viral RNAs reflect the array of polypeptides in the EcoRI J and N fragments (map positions 81.2 to 85.0) of the Autographa californica nuclear polyhedrosis virus genome. **J. Virol.**, v. 61, n. 10, p. 3048-3057, 1987.

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. Genome of the most widely used viral biopesticide: Anticarsia gemmatalis multiple nucleopolyhedrovirus. **J. Gen. Virol.**, v. 87, n. Pt 11, p. 3233-3250, 2006.

OLIVEIRA, J. V. C. **Expressão temporal dos genes do nucleopoliedrovírus Anticarsia gemmatalis e sua influência sobre a célula**. 2010. 243 (Doutorado). Programa de Pós-Graduação Interunidades em Biotecnologia USP/Instituto Butantan/IPT, Universidade de São Paulo, São Paulo.

OOI, B. G.; MILLER, L. K. Regulation of host RNA levels during baculovirus infection. **Virology**, v. 166, n. 2, p. 515-523, 1988.

PANG, Y.; YU, J.; WANG, L.; HU, X.; BAO, W.; LI, G.; CHEN, C.; HAN, H.; HU, S.; YANG, H. Sequence analysis of the Spodoptera litura multicapsid nucleopolyhedrovirus genome. **Virology**, v. 287, n. 2, p. 391-404, 2001.

PARK, E. J.; YIN, C. M.; BURAND, J. P. Baculovirus replication alters hormone-regulated host development. J. Gen. Virol., v. 77 (Pt 3), p. 547-554, 1996.

PASHLEY, D. P.; JOHNSON, S. J. Genetic population structure of migratory moths: the velvetbean caterpillar Lepidoptera: Noctuidae. **Ann. Entomol. Soc. Am.,** v. 79, n. 1, p. 26-30, 1986.

PASSARELLI, A. L.; MILLER, L. K. Three baculovirus genes involved in late and very late gene expression: ie-1, ie-n, and lef-2. **J. Virol.**, v. 67, n. 4, p. 2149-2158, 1993.

_____. In vivo and in vitro analyses of recombinant baculoviruses lacking a functional cg30 gene. **J. Virol.,** v. 68, n. 2, p. 1186-1190, 1994.

POSSEE, R. D.; ROHRMANN, G. F. Baculovirus genome organization and evolution. In: MILLER, L. K. (Ed.). **The baculoviruses**. New York: Plenum, 1997. p.109–140.

PRIKHOD'KO, E. A.; LU, A.; WILSON, J. A.; MILLER, L. K. In vivo and in vitro analysis of baculovirus ie-2 mutants. J. Virol., v. 73, n. 3, p. 2460-2468, 1999.

PRIKHOD'KO, E. A.; MILLER, L. K. The baculovirus PE38 protein augments apoptosis induced by transactivator IE1. J. Virol., v. 73, n. 8, p. 6691-6699, 1999.

RANCUREL, C.; KHOSRAVI, M.; DUNKER, A. K.; ROMERO, P. R.; KARLIN, D. Overlapping genes produce proteins with unusual sequence properties and offer insight into de novo protein creation. **J. Virol.**, v. 83, n. 20, p. 10719-10736, 2009.

RODEMS, S. M.; FRIESEN, P. D. The hr5 transcriptional enhancer stimulates early expression from the Autographa californica nuclear polyhedrosis virus genome but is not required for virus replication. **J. Virol.**, v. 67, n. 10, p. 5776-5785, 1993.

_____. Transcriptional enhancer activity of hr5 requires dual-palindrome half sites that mediate binding of a dimeric form of the baculovirus transregulator IE1. **J. Virol.**, v. 69, n. 9, p. 5368-5375, 1995.

ROHRMANN, G. F. Polyhedrin structure. J. Gen. Virol., v. 67 (Pt 8), p. 1499-1513, 1986.

_____. Baculovirus structural proteins. **J. Gen. Virol.**, v. 73 (Pt 4), p. 749-761, 1992.

_____. Baculovirus Molecular Biology. Bethesda, 2011. Disponível em: < <u>http://www.ncbi.nlm.nih.gov/books/NBK49500/</u> >. Acesso em: 03 mar. 2012.

RUTHERFORD, K.; PARKHILL, J.; CROOK, J.; HORSNELL, T.; RICE, P.; RAJANDREAM, M. A.; BARRELL, B. Artemis: sequence visualization and annotation. **Bioinformatics**, v. 16, n. 10, p. 944-945, 2000.

SAURIN, A. J.; BORDEN, K. L.; BODDY, M. N.; FREEMONT, P. S. Does this have a familiar RING? **Trends Biochem. Sci.**, v. 21, n. 6, p. 208-214, 1996.

SHACKELTON, L. A.; HOLMES, E. C. The evolution of large DNA viruses: combining genomic information of viruses and their hosts. **Trends Microbiol.**, v. 12, n. 10, p. 458-465, 2004.

SIMON, O.; WILLIAMS, T.; LOPEZ-FERBER, M.; CABALLERO, P. Genetic structure of a Spodoptera frugiperda nucleopolyhedrovirus population: high prevalence of deletion genotypes. **Appl. Environ. Microbiol.,** v. 70, n. 9, p. 5579-5588, 2004.

SMITH, G. E.; SUMMERS, M. D. Restriction Map of Rachiplusia ou and Rachiplusia ou-Autographa californica Baculovirus Recombinants. **J. Virol.**, v. 33, n. 1, p. 311-319, 1980.

SMITH, I. Misleading messengers? Interpreting baculovirus transcriptional array profiles. J. Virol., v. 81, n. 14, p. 7819-7820, 2007.

SOARES, J. S.; RIBEIRO, B. M. Pathology of Anticarsia gemmatalis larvae infected by two recombinant A. gemmatalis multicapsid nucleopolyhedroviruses. **Res. Microbiol.,** v. 156, n. 2, p. 263-269, 2005.

SONNHAMMER, E. L.; DURBIN, R. A dot-matrix program with dynamic threshold control suited for genomic DNA and protein sequence analysis. **Gene**, v. 167, n. 1-2, p. GC1-10, 1995.

SOSA-GÓMEZ, D. R. Intraspecific variation and population structure of the Velvetbean Caterpillar, Anticarsia gemmatalis Hübner, 1818 (Insecta: Lepidoptera: Noctuidae). **Genet. Mol. Biol.**, v. 27, n. 3, p. 378-384, 2004.

STEINHAUS, E. A.; MARSH, G. A. Report of diagnoses of diseased insects 1951-1961. Hilgardia v. 33, p. 349-490, 1962.

TAJIMA, F. Measurement of DNA polymorphism. In: TAKAHATA, N. e CLARK, A. G. (Ed.). **Mechanisms of molecular evolution**. Sunderland, MA: Sinauer Associates. Inc., 1993. p.37-59.

TAUTZ, D.; DOMAZET-LOSO, T. The evolutionary origin of orphan genes. Nat. Rev. Genet., v. 12, n. 10, p. 692-702, 2011.

THIEM, S. M.; CHENG, X. W. Baculovirus Host-Range. Virol. Sin., v. 24, n. 5, p. 436-457, 2009.

THOMPSON, C. G.; SCOTT, D. W.; WICKMAN, B. E. Long-term persistence of the nuclear polyhedrosis virus of the Douglas-fir tussock moth, Orgyia pseudotsugata (Lepidoptera: Lymantriidae), in forest soil. **Environ. Entomol.**, v. 10, p. 254–255, 1981.

THOMPSON, J. D.; HIGGINS, D. G.; GIBSON, T. J. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. **Nucleic Acids Res.**, v. 22, n. 22, p. 4673-4680, 1994.

TRIFINOV, E. N. The multiple codes of nucleotide sequences. Bull. Math. Biol., v. 51, p. 417-432, 1989.

VAN OERS, M. M.; ABMA-HENKENS, M. H.; HERNIOU, E. A.; DE GROOT, J. C.; PETERS, S.; VLAK, J. M. Genome sequence of Chrysodeixis chalcites nucleopolyhedrovirus, a baculovirus with two DNA photolyase genes. J. Gen. Virol., v. 86, n. Pt 7, p. 2069-2080, 2005.

VAN OERS, M. M.; VLAK, J. M. Baculovirus genomics. Curr. Drug Targets, v. 8, n. 10, p. 1051-1068, 2007.

VAN STRIEN, E. A.; FAKTOR, O.; HU, Z. H.; ZUIDEMA, D.; GOLDBACH, R. W.; VLAK, J. M. Baculoviruses contain a gene for the large subunit of ribonucleotide reductase. **J. Gen. Virol.**, v. 78 (Pt 9), p. 2365-2377, 1997.

VLAK, J. M.; SMITH, G. E. Orientation of the Genome of Autographa californica Nuclear Polyhedrosis Virus: a Proposal. J. Virol., v. 41, n. 3, p. 1118-1121, 1982.

VOLKMAN, L. E. Baculovirus infectivity and the actin cytoskeleton. **Curr. Drug Targets**, v. 8, n. 10, p. 1075-1083, 2007.

VOLKMAN, L. E.; BLISSARD, G. W.; FRIESEN, P. D.; KEDDIE, B. A.; POSSEE, R. D.; THEILMANN, D. A. Baculoviridae. In: MURPHY, F. A.;FAUQUET, C. M.;BISHOP, D. H. L.;GHABRIAL, S. A.;JARVIS, A. W.;MARTELLI, G. P.;MAYO, M. A. e SUMMERS, M. D. (Ed.). **Virus taxonomy**. New York: Springer-Verlag, 1995. p.104-113.

WATTERSON, G. A. On the number of segregating sites in genetical models without recombination. **Theor. Popul. Biol.**, v. 7, n. 2, p. 256-276, 1975.

WERTZ, G. W.; PEREPELITSA, V. P.; BALL, L. A. Gene rearrangement attenuates expression and lethality of a nonsegmented negative strand RNA virus. **Proc. Natl. Acad.** Sci. USA, v. 95, n. 7, p. 3501-3506, 1998.

WHITT, M. A.; MANNING, J. S. Role of chelating agents, monovalent anion and cation in the dissociation of Autograph californicu nuclear polyhedrosis virus occlusion body matrix by zinc chloride. **J. Inverteb. Pathol.**, v. 49, p. 61-69, 1987.

WILLIAMS, B. A.; SLAMOVITS, C. H.; PATRON, N. J.; FAST, N. M.; KEELING, P. J. A high frequency of overlapping gene expression in compacted eukaryotic genomes. **Proc. Natl.** Acad. Sci. USA, v. 102, n. 31, p. 10936-10941, 2005.

WILLIS, L. G.; SEIPP, R.; STEWART, T. M.; ERLANDSON, M. A.; THEILMANN, D. A. Sequence analysis of the complete genome of Trichoplusia ni single nucleopolyhedrovirus and the identification of a baculoviral photolyase gene. **Virology**, v. 338, n. 2, p. 209-226, 2005.

WORMLEATON, S.; KUZIO, J.; WINSTANLEY, D. The complete sequence of the Adoxophyes orana granulovirus genome. **Virology**, v. 311, n. 2, p. 350-365, 2003.

WU, L. C.; WANG, Z. W.; TSAN, J. T.; SPILLMAN, M. A.; PHUNG, A.; XU, X. L.; YANG, M. C.; HWANG, L. Y.; BOWCOCK, A. M.; BAER, R. Identification of a RING protein that can interact in vivo with the BRCA1 gene product. **Nat. Genet.**, v. 14, n. 4, p. 430-440, 1996.

XING, K.; DENG, R.; WANG, J.; FENG, J.; HUANG, M.; WANG, X. Analysis and prediction of baculovirus promoter sequences. **Virus Res.**, v. 113, n. 1, p. 64-71, 2005.

YANAI, I.; WOLF, Y. I.; KOONIN, E. V. Evolution of gene fusions: horizontal transfer versus independent events. **Genome Biol.**, v. 3, n. 5, p. research0024, 2002.

YOUNG, S. Y.; YEARIAN, W. C. Movement of a nuclear polyhedrosis virus from soil to soybean and transmission in Anticarsia gemmatalis (Hübner) (Lepidoptera: Noctuidae) populations on soybean. **Environ. Entomol.**, v. 15, p. 573–580, 1986.

ZAGORDI, O.; GEYRHOFER, L.; ROTH, V.; BEERENWINKEL, N. Deep sequencing of a genetically heterogeneous sample: local haplotype reconstruction and read error correction. **J. Comput. Biol.**, v. 17, n. 3, p. 417-428, 2010.

ZANOTTO, P. M.; KESSING, B. D.; MARUNIAK, J. E. Phylogenetic interrelationships among baculoviruses: evolutionary rates and host associations. **J. Invertebr. Pathol.,** v. 62, n. 2, p. 147-164, 1993.

ZANOTTO, P. M. A.; KRAKAUER, D. C. Complete genome viral phylogenies suggests the concerted evolution of regulatory cores and accessory satellites. **PLoS One**, v. 3, n. 10, p. e3500, 2008.

ZEMSKOV, E. A.; KANG, W.; MAEDA, S. Evidence for nucleic acid binding ability and nucleosome association of Bombyx mori nucleopolyhedrovirus BRO proteins. J. Virol., v. 74, n. 15, p. 6784-6789, 2000.

ZHU, S. Y.; YI, J. P.; SHEN, W. D.; WANG, L. Q.; HE, H. G.; WANG, Y.; LI, B.; WANG, W. B. Genomic sequence, organization and characteristics of a new nucleopolyhedrovirus isolated from Clanis bilineata larva. **BMC Genomics**, v. 10, p. 91, 2009.

ANEXO A – Tabelas Suplementares

Tabela A1 – Matriz o	le presença e ausência das ORFs de AgMNPV-2D nos 17 isolados geográficos. 0 e 1
indicam	respectivamente, presença e ausência. Em negrito (e rosa) estão destacadas as ORFs
não com	partilhadas por todos os isolados. As cores apresentadas referem-se as da Figura 43.

ORF	Gene	2D	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	42	43	Total
001	polh	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	18
002	Ag2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	18
003	lef2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	18
004	Ag4	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	18
005	Ag5	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	18
006	bro-a	1	1	0	0	0	0	1	0	0	0	1	0	0	0	0	1	0	1	6
007	bro-b	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	18
008	ptp2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	18
009	ptp1/bvp	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	18
010	Ag10	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	18
011	bro-c	1	1	1	1	1	1	1	0	1	1	0	0	1	1	1	1	1	1	15
012	pif2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	18
013	arif1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	18
014	Ag14	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	18
015	Ag15	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	18
016	Ag16	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	18
017	odve26	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	18
018	egt	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	18
019	lef1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	18
020	Ag20	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	18
021	Ag21	1	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	3
022	F-protein	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	18
023	Ag23	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	18
024	lef11	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	18
025	39K/pp31	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	18
026	ubi	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	18
027	Ag27	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	18
028	hisP	1	0	1	0	0	0	1	1	1	1	1	0	1	0	0	1	0	1	10
029	fgf	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	18
030	ctl-2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	18
031	parp-like	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	18
032	sod	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	18
033	bro-d	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	18
034	iap3	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	18
035	Ag35	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	18
036	Ag36	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	18
037	Ag37	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	18
038	Ag38	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	18
039	lef6	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	18
040	iap1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	18
041	Ag41	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	18
042	dbp	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	18
043	pkip1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	18
044	p47	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	18
045	lef12	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	18
046	gta	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	18
047	Ag47	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	Ţ	. 18
																			(con	itinua)

048 Ag48 1 <th></th> <th>(co</th> <th>ntinu</th> <th>ıação)</th>																			(co	ntinu	ıação)
049 odve66 1<	048	Ag48	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	18
$ \begin{array}{c c c c c c c c c c c c c c c c c c c $	049	odve66	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	18
$ \begin{array}{c c c c c c c c c c c c c c c c c c c $	050	etm	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	18
$ \begin{array}{cccccccccccccccccccccccccccccccccccc$	051	lef8	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	18
$\begin{array}{cccccccccccccccccccccccccccccccccccc$	052	Ag52	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	18
$\begin{array}{cccccccccccccccccccccccccccccccccccc$	053	Ag53	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	18
$ \begin{array}{cccccccccccccccccccccccccccccccccccc$	054	1ef10	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	18
$ \begin{array}{cccccccccccccccccccccccccccccccccccc$	054	vn1054	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	18
$ \begin{array}{cccccccccccccccccccccccccccccccccccc$	055	A a 56	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	18
$ \begin{array}{c} 0.7 \\ Ag S \\ Bi \\ I \\ $	057	Ag57	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	18
$ \begin{array}{cccccccccccccccccccccccccccccccccccc$	058	Ag58	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	18
$ \begin{array}{cccccccccccccccccccccccccccccccccccc$	050	Ag50	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	10
$ \begin{array}{cccccccccccccccccccccccccccccccccccc$	059	Ag59	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	10
$ \begin{array}{cccccccccccccccccccccccccccccccccccc$	000	Ag00	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	10
$ \begin{array}{cccccccccccccccccccccccccccccccccccc$	061	23KJP	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	10
$ \begin{array}{cccccccccccccccccccccccccccccccccccc$	062		1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	18
$ \begin{array}{cccccccccccccccccccccccccccccccccccc$	005	Dro-e	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	10
$\begin{array}{cccccccccccccccccccccccccccccccccccc$	064	Ag04	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	3
$\begin{array}{cccccccccccccccccccccccccccccccccccc$	065	DNApol	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	18
$\begin{array}{c c c c c c c c c c c c c c c c c c c $	066	desmoplakin	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	l	1	1	18
$\begin{array}{c c c c c c c c c c c c c c c c c c c $	06/	lef3	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	18
$\begin{array}{c c c c c c c c c c c c c c c c c c c $	068	pif6	1	1	1	1	l	1	1	1	1	1	1	1	1	1	1	l	I	1	18
$ \begin{array}{cccccccccccccccccccccccccccccccccccc$	069	met	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	18
$ \begin{array}{cccccccccccccccccccccccccccccccccccc$	070	iap2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	18
$ \begin{array}{cccccccccccccccccccccccccccccccccccc$	071	Ag71	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	18
$ \begin{array}{cccccccccccccccccccccccccccccccccccc$	072	Ag72	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	18
$ \begin{array}{cccccccccccccccccccccccccccccccccccc$	073	Ag73	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	18
$ \begin{array}{cccccccccccccccccccccccccccccccccccc$	074	Ag74	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	18
$\begin{array}{c c c c c c c c c c c c c c c c c c c $	075	Ag75	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	18
$\begin{array}{c c c c c c c c c c c c c c c c c c c $	076	vlf1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	18
$\begin{array}{c c c c c c c c c c c c c c c c c c c $	077	Ag77	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	18
$\begin{array}{cccccccccccccccccccccccccccccccccccc$	078	A g78	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	18
$\begin{array}{c c c c c c c c c c c c c c c c c c c $	079	gp41	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	18
$\begin{array}{c c c c c c c c c c c c c c c c c c c $	080	Ag80	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	18
$\begin{array}{c c c c c c c c c c c c c c c c c c c $	081	tlp20	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	18
$\begin{array}{cccccccccccccccccccccccccccccccccccc$	082	vp91/p95	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	18
$\begin{array}{c c c c c c c c c c c c c c c c c c c $	083	ag83	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	18
$\begin{array}{c ccccccccccccccccccccccccccccccccccc$	084	p 15	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	18
$\begin{array}{c ccccccccccccccccccccccccccccccccccc$	085	cg30	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	18
$\begin{array}{c c c c c c c c c c c c c c c c c c c $	086	vp39	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	18
$\begin{array}{c c c c c c c c c c c c c c c c c c c $	087	lef4	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	18
$\begin{array}{c c c c c c c c c c c c c c c c c c c $	088	Ag88	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	18
$\begin{array}{c ccccccccccccccccccccccccccccccccccc$	089	p33	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	18
$\begin{array}{c ccccccccccccccccccccccccccccccccccc$	090	p18	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	18
$\begin{array}{c c c c c c c c c c c c c c c c c c c $	091	odve25	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	18
$\begin{array}{c ccccccccccccccccccccccccccccccccccc$	092	helicase	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	18
$\begin{array}{c ccccccccccccccccccccccccccccccccccc$	093	nif4	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	18
$\begin{array}{c ccccccccccccccccccccccccccccccccccc$	094	38k	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	18
$\begin{array}{c ccccccccccccccccccccccccccccccccccc$	095	lef5	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	18
$\begin{array}{c ccccccccccccccccccccccccccccccccccc$	096	<i>p</i> 6.9	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	18
$\begin{array}{c ccccccccccccccccccccccccccccccccccc$	097	p40/c42	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	18
$\begin{array}{c ccccccccccccccccccccccccccccccccccc$	098	p12	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	18
$100 \blacksquare p87/vp80 \qquad 1 \ 1 \ 1 \ 1 \ 1 \ 1 \ 1 \ 1 \ 1 \ 1$	099	p48	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	18
(continua)	100	p = p + 0 p = p + 0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	18
		r	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	(con	tinua)

																		(0	conc	lusão)
101	he65a	1	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	1	4
102	he65b	1	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	1	4
103	pnk/pnl	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	18
104	Ag104	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	18
105	bro-f	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	18
106	Ag106	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	18
107	odv-ec43	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	18
108	Ag108	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	18
109	Ag109	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	18
110	Ag110	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	18
111	Ag111	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	18
112	Ag112	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	18
113	Ag113	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	18
114	pif3	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	18
115	bro-g	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	18
116	A g116	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	18
117	p if1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	18
118	A g118	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	18
119	A g119	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	18
120	A g120	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	18
121	Ag121	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	18
122	<i>v-trex</i>	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	18
123	lef7	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	18
124	g p64	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	18
125	p 24	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	18
126	g p16	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	18
127	p p34/pep	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	18
128	p 25	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	18
129	alkexo	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	18
130	Ag130	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	18
131	Ag131	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	18
132	p 26	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	18
133	p 10	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	18
134	p 74	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	18
135	bro-h	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	18
136	m e53	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	18
137	📕 ie0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	18
138	p 49	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	18
139	dve18	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	18
140	odve27	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	18
141	Ag141	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	18
142	Ag142	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	18
143	ie1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	18
144	odve56/nif5	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	18
145	ie2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	18
146	Ag146	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	18
140	A 0147	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	18
148	ne-38-like-a	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
140	pe=30 inte-u	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
150	= pe-30-nke-0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	18
150	nk1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	18
151	= pr1 = 1620cansid	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	10
132	1 029cupsid	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	10

		estão	dest	acac	las	as	ORF	Fs n	ão	com	parti	ilhac	las	por	tod	os	os i	sola	dos.	As	cores
		aprese	entad	as c	orre	spon	dem	as o	la F	igura	a 43.										
ORF		Gene	2D	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	42	43	Total
011a		bro-c1	0	0	0	0	0	0	0	1	0	0	1	1	0	0	0	0	0	0	3
011b		bro-c2	0	0	0	0	0	0	0	1	0	0	1	1	0	0	0	0	0	0	3
021a		Ag21a	0	1	1	1	1	1	1	1	1	1	1	0	0	1	1	1	1	1	15
021b		Ag21b	0	1	1	1	1	1	1	1	1	1	1	0	0	1	1	1	1	1	15
028b		hisP2	0	1	0	1	1	1	0	0	0	0	0	1	0	1	1	0	1	0	8
064b		Ag64b	0	1	1	1	1	1	1	0	1	1	1	1	0	1	1	1	1	1	15
101a		he65	0	1	1	1	1	1	1	1	1	1	1	1	0	1	0	1	1	0	14
148a		pe-38	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	17
025b		Ag25b	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	2
029b		Ag29b	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	18
051b		Ag51b	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	18
001b		Ag1b	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	18
002b		Ag2b	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
004b		Ag4b	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	18
009b		Ag9b	0	1	1	1	1	0	1	1	1	1	1	1	0	0	1	1	1	1	14
011c		Agllc	1	1	0	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	16
021d		Ag21d	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	18
031b		Ag31b	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	18
050b		Ag50b	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	18
102b		Ag102b	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	18
104b	_	Ag104b	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	18
117c	_	Ag117c	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	18
120b		Ag120b	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	18
123c	_	Ag123c	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	18
130b	_	Ag130b	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	18
136b	_	Ag136b	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	18
136c	_	Ag136c	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	18
021c	-	Ag21c	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	18
022b	-	Ag22b	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	18
052b	-	Ag52b	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	17
113b	-	Ag113b	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	18
116b		Ag116b	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	18
116c		Ag116c	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	18
117b		Ag117b	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	18
123b		Ag123b	1	1	1	1	1	1	1	1	1	1	1	1	l	1	1	1	1	1	18
124b		Ag124b	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	17
147b		Ag147b	0	0	1	0	1	0	1	0	0	0	1	0	0	0	0	1	1	1	7

Tabela A2 – Matriz de presença e ausência das ORFs de supostos genes dos 17 isolados geográficos e no AgMNPV-2D. 0 e 1 indicam, respectivamente, presença e ausência. Em **negrito** estão destacadas as ORFs não compartilhadas por todos os isolados. As cores apresentadas correspondem as da Figura 43.

ANEXO B – Artigo

Infection, Genetics and Evolution 18 (2013) 202-212

Contents lists available at SciVerse ScienceDirect

Infection, Genetics and Evolution

journal homepage: www.elsevier.com/locate/meegid

Genomic mosaicism in two strains of dengue virus type 3

Christian Julián Villabona-Arenas, Anderson Fernandes de Brito, Paolo Marinho de Andrade Zanotto*

Laboratory of Molecular Evolution and Bioinformatics, Department of Microbiology, Institute of Biomedical Sciences, University of São Paulo, Av. Prof. Lineu Prestes, 1734, CEP: 05508-000, São Paulo, SP, Brazil

ARTICLE INFO

Article history: Received 5 March 2013 Received in revised form 14 May 2013 Accepted 15 May 2013 Available online 30 May 2013

Keywords: Dengue virus type 3 Phylogeny Recombination Split decomposition

1. Introduction

The availability of increasing sequence information from different localities and refined phylogenetic tools make it possible to address many of the processes that lead to genetic diversification, adaptation and spread of *Dengue virus* (DENV) (Weaver and Vasilakis, 2009). Populations of RNA viruses usually harbor genetic diversity, associated with high mutation rates, small effective population and large population sizes (Simon-Loriere and Holmes, 2011). The exchange of genetic material is an important mechanism that allows adaptive change among natural populations and evidences its ubiquitous role in evolution (Awadalla, 2003).

Dengue fever in humans is caused by enveloped, positive single stranded RNA ([+] ssRNA) mosquito-borne viruses, which belong to the genus *Flavivirus* (Family: Flaviviridae) and comprise four sero-types (DENV-1 to DENV-4) (Karabatsos, 1985). Their genomes (~11 kbp) encode a single open reading frame, flanked by highly structured 5' and 3' untranslated regions (UTRs) (Lindenbach and Rice, 2003; Rice et al., 1985). The N-terminal of the polyprotein encodes the three structural proteins (C-prM-E), followed by seven non-structural (NS) proteins (NS1–NS2A–NS2B–NS3-NS4A–NS4B–NS5) (Chambers et al., 1990; Rice et al., 1985).

Concomitant infection with two different strains (a prerequisite for DENV recombination) was demonstrated both in humans and in mosquitoes (Bharaj et al., 2008; Loroño-Pino et al., 1999; Laille et al., 1991; Montenegro et al., 2006; Wang et al., 2003). Recently, Thai et al. (2012) documented distinct lineages within the same host. Evidence of recombination between diverse dengue strains

* Corresponding author. Tel.: +55 11 3091 8453.

E-mail address: pzanotto@usp.br (P.M. de Andrade Zanotto).

ABSTRACT

Recombination is a significant factor driving genomic evolution, but it is not well understood in Dengue virus. We used phylogenetic methods to search for recombination in 636 Dengue virus type 3 (DENV-3) genomes and unveiled complex recombination patterns in two strains, which appear to be the outcome of recombination between genotype II and genotype I parental DENV-3 lineages. Our findings of genomic mosaic structures suggest that strand switching during RNA synthesis may be involved in the generation of genetic diversity in dengue viruses.

© 2013 Elsevier B.V. All rights reserved.

has arisen (Aaskov et al., 2007; AbuBakar et al., 2002; Carvalho et al., 2010; Chen et al., 2008; Craig et al., 2003; Lavergne et al., 2006; Uzcategui et al., 2001; Worobey et al., 1999) since the first report of recombination in natural populations in 1999 (Holmes et al., 1999). Also, recombination was shown in other [+] ssRNA viruses, (Becher and Tautz, 2011; Carney et al., 2011; Chuang and Chen, 2009; Gould et al., 2004; Kalinina et al., 2002; Reiter et al., 2011), Retroviruses (Nájera et al., 2002) and [-] ssRNA viruses (Sibold et al., 1999). Two main mechanisms have been proposed for viral RNA recombination: (i) non-replicative breakage and rejoining and (ii) replicative template switching (Cooper, 1974; Nagy and Simon, 1997). The first model is a viral replication-independent RNA recombination process in which RNAs spontaneously rearrange their sequences, as found in tissue culture cells infected by polioviruses (Gmyl et al., 1999) and pestiviruses (Becher and Tautz, 2011). The second mechanism predicts that during replication the viral RdRp pauses RNA synthesis and eventually jumps, together with the incomplete nascent RNA, to another template in order to continue RNA synthesis. This mechanism has also been studied for [+] RNA such as the Brome mosaic virus and polioviruses, pestiviruses, carmoviruses and tombusviruses (Arnold and Cameron, 1999; Cheng and Nagy, 2003; Dahourou et al., 2002; Duggal et al., 1997; Simon and Nagy, 1996).

Due to modern transportation, wide-range dispersion of viruses can take place within 24–36 h resulting in several epidemic outbreaks, some of which involve the co-circulation of different viral strains, which increases the likelihood of mixed infections (Gould and Gritsun, 2006). Herein, we made use of available data and distinct analytical approaches to identify for the first time recombinant viruses in whole-genome alignments of DENV-3.





CrossMark

^{1567-1348/\$ -} see front matter @ 2013 Elsevier B.V. All rights reserved. http://dx.doi.org/10.1016/j.meegid.2013.05.012

2. Materials and methods

2.1. Data set

Genomic DENV-3 strains were retrieved from GenBank (Table S1, in the supplemental material) and their source feature inspected, as a caution required in obtaining data from a public database, to check each entry name and to exclude mutants and non-human/non-mosquitoes isolates. Alignments were built with the Muscle 3.8.31 program (Edgar, 2004a,b), manually inspected, and trimmed to include only coding regions.

2.2. Detecting recombination

We used the automated algorithms contained within the recombination detection program 3.0 (RDP3) (Martin et al., 2010) to screen for putative recombination events in the coding region alignment. We used the default settings for each algorithm with a highest acceptable *p*-value of 0.01 and the standard Bonferroni correction (a *p*-value modifier that decrease the *p*-value cutoff according to the size of the dataset being examined) and set the analysis to report all hits detected by 6 or more algorithms. Once a putative recombination event was detected, a smaller dataset was built from selected sequences that included the likely recombinant, surrogate parental taxa, and reference strains.

2.3. Detection of recombinant breakpoints in the whole-genome alignment

Two Hidden Markov Models procedures (HMM) were used to corroborate the results from the automated tests and accurately locate breakpoints: (i) The Husmeier and Wright approach (Husmeier and McGuire, 2003) with four taxa (a likely recombinant, the surrogate parental taxa and one reference strain) and a timereversible F84 substitution model (Felsenstein and Churchill, 1995) implemented in the Topali 2.5 software (Milne et al., 2009); and (ii) the Westesson and Holmes approach (86) that follows the structural EM heuristic (Friedman et al., 2002) using the RecHMM software by setting the minimum cutoff length for a recombinant region to 400 bp and using two additional reference strains. Significant tree topology changes allow partition of the alignment into genomic regions with apparent distinct evolutionary histories. For instance, for four taxa, there are three different unrooted tree topologies; in the presence of recombination, the tree topology can change and thus become a random variable that depends on the sites under consideration (the hidden states of the HMM represent the three topologies).

2.4. Phylogenetic network analyses

Alignments were built for each detected recombinant, including the most plausible surrogate parental taxa and three reference sequences from each genotype (free of recombination signal as evidenced by further automated detection tests). A pairwise homoplasy index (PHI) based on the principle of refined incompatibility was calculated for each alignment to test for recombination (Bruen et al., 2006). Afterwards, Split decomposition analyses (Bandelt and Dress, 1992) were done to assess the presence of a 'phylogenetic network' as an indication of reticulate evolution. This approach enhances phylogenetic analyses of distance data by unveiling incompatible groupings of taxa (splits) that arise when recombination and/or horizontal gene transfer have occurred. Analyses were done with the SplitsTree 4.0 software (Huson and Kloepper, 2005), with a bootstrapping of 100 replicates, using a general time reversible substitution model (GTR). The BEAST software package (Drummond et al., 2012) was used to estimate the proportion of invariable sites (I) among site rate variation (γ); nucleotide frequencies; and substitution rates for the model.

2.5. Phylogenetic reconstruction

A Maximum Clade Credibility tree (MCC) and divergence times for each alignment were inferred by using a Bayesian Markov chain Monte Carlo (MCMC) approach as implemented in BEAST. The estimated GTR + $I + \gamma$ model was used together with a relaxed molecular clock (Drummond et al., 2006) and a Bayesian Skyline population growth model (Drummond et al., 2005). Four MCMC chain lengths were run for 80 million generations sampling every 8000th generation. Four independent runs were done to improve convergence and effective sample sizes were assessed using Tracer 1.4 (http://beast.bio.ed.ac.uk/Tracer).

3. Results

3.1. DENV-3 recombinants

We obtained 636 genomic sequences from GenBank (Table S1, in the supplemental material) and the final alignment included 10,170 coding nucleotide positions. We detected a significant recombination event, with an exceedingly small probability of 5.7E-18 of random occurrence obtained with multiple comparisons correction by all the algorithms implemented in RDP 3.0 (RDP, GENECONV, BootScan, MaxChi, Chimaera, Siscan and 3Seq). For this single event (Fig. 1) DQ401690 (Indonesia 1982) and AY496879 (Philippines 1997) appeared as recombinants, while AY858038 (Indonesia 1988), AY496877 (Bangladesh 2002) and FN429900 (Malaysia 1999) appeared as surrogate parental taxa.

3.2. Phylogenetic network of recombinants

We sought to characterize these events using different criteria and the dataset from Table 1. We used a phylogenetic network approach to provide an implicit picture of evolutionary relationships where the sets of parallel edges (splits) represented incompatible and ambiguous signals in the dataset. In this analysis, we included the complete monophyletic groups to which the surrogate parental taxa belong. These groups comprised the surrogate parental lineage and were retrieved from a maximum likelihood tree inferred with Garli 2.0 (Zwickl, 2006) and the 636 wholegenome sequences (Fig. S1). Our findings provided statistically significant evidence for recombination as indicated by a PHI test (p-value <0.05) and the split decomposition output showed multiple bands of parallel edges associating PH-1997 and ID-1982 with two different genotypes (Fig. 2A). Nevertheless, these split networks did not conclusively support genotype assignment either to genotype II or I, but suggested that there was evidence for both.

We created a chimera to better scrutinize our findings in a computer simulation, since the comparative data did not portray a strong tree-like structure. It was built *in silico* by joining together genomic regions from DENV-1 and DENV-4 serotypes and using it alongside other DENV sequences in a Split Decomposition analysis (Fig. S1, in the supplemental material). Fig. S2A illustrates the resulting phylogenetic network. The interesting result was that the chimera had a region of split decomposition due to conflicting evolutionary histories involving two diverse ancestral monophyletic groups, as in our real data.



Fig. 1. Results from the automated tests with RDP. The recombinants ID-1982 and PH-1997 were aligned to a schematic representation of the DENV genome positioned at the center. For each recombinant, the internal rectangles indicate the surrogate parental taxa (labeled as in Table 1).

Table 1

Sequences used in this study following the automated detection test results conducted with RDP.

Genbank Acc. no.	Local and year of isolation	Label	Description
AY744678	French Polynesia, 1989	PF-1989	Genotype I
FN429904	Malaysia, 2000	MY-2000	Genotype I
EU081221	Singapore, 2005	SG-2005	Genotype I
AY676353	Thailand, 1987	TH-1987	Genotype II
GU131906	Cambodia, 2003	KH-2003	Genotype II
FJ461337	Vietnam, 2008	VN-2008	Genotype II
JQ411814	Sri Lanka, 1989	LK-1989	Genotype III
FJ898459	Trinidad and Tobago, 2002	TT-2002	Genotype III
GU131964	Colombia, 2006	CO-2006	Genotype III
DQ401690	Indonesia, 1982	ID-1982	Recombinant
AY496879	Philippines, 1997	PH-1997	Recombinant
AY496877	Bangladesh, 2002	BD-2002	Minor parent ^a
AY496871	Bangladesh, 2002	BD-2002a	
AY496873	Bangladesh, 2002	BD-2002b	
AY496874	Bangladesh, 2002	BD-2002c	
FN429913	Malaysia, 2004	MY-2004	
AY858038	Indonesia, 1988	ID-1988	Major parent ^a
FN429899	Malaysia, 1997	MY-1997	
FN429900	Malaysia, 1999	MY-1999	

^a Monophyletic groups.

3.3. Recombination breakpoint locations

We predicted accurately the recombination breakpoints using algorithms for estimating both tree structure and hidden Markov model parameters (HMM) (Husmeier and McGuire, 2003). Fig. 3A and B shows the corresponding probabilistic graphical model applied to four- and six-taxa ID-1982 alignments respectively (Fig. 4A and B for PH-1997 datasets). Most predictions with each HMM methodology were in agreement and confirmed the existence of multiple breakpoints.

Notably, HMM approaches predicted more breakpoints than RDP (Fig. 1), some of which matched gene boundaries except for the NS5 gene, which encompassed a mosaic region. Within this gene, the RdRp motifs were conserved, except for motif P that had a polymorphic residue in a codon (position 3293 Arg \rightarrow Trp) involved in priming loop stability (88). Interestingly, the only three

sequences that had this amino acid replacement corresponded to the two recombinants and one surrogate parental virus (BD-2002) (Fig. S3, in the supplemental material).

3.4. Effect of recombination on the accuracy of phylogeny

We calculated PHI, applied the network approach and inferred Maximum clade credibility trees using Bayesian methods to study the phylogenetic congruence among different regions. We selected the partitions for each recombinant following Topali's predictions and BD-2002 and ID-1988 as parents for simplicity. We also limited the number of regions to six because of the multiple breakpoints within the NS5 gene in the case of PH-1997. Figs. 3 and 4C networks evidenced that different regions showed a different evolutionary history that resulted in inconclusive genotyping. In contrast, topologies obtained using a dataset free of recombination signal were not reticulated and no taxon was ambiguously genotyped (Fig. S4, in the supplemental material). PHI p-values did not indicate conflicting evolutionary histories (p-values >0.05) except for the region c of the ID-1982 dataset, where we obtained statistically significant evidence for recombination (p-value of 7.6E-4). This could be a consequence of the topology probabilities around nucleotide position 7470 (Fig. 4A) that did not discriminate between topologies I and III.

Figs. 5 and 6 show the MCC trees for genomes and regions. Taxa associations resembled those of the phylogenetic network analyses, where both recombinants had different genotype assignment: some regions grouped in Genotype II and that others grouped in Genotype I. The TMRCA for regions *b*, *d* and *f* (where BD-2002 and ID-1982 grouped together) were oddly higher (116.9, 213.4 and 364.1 years, respectively) compared with that estimated for the genome (66.1 years) and the remaining regions (μ = 57.60, σ = 3.42) (Fig. 5). Recombination events are known to introduce bias when timing events in a viral genealogy (Awadalla, 2003), but it seems that it was the coalescent method that had problems when taxa sampled 20 years apart grouped together. In contrast, PH-1997 and BD-2002, which were sampled at a shorter time interval, did not show overestimated TMRCAs values when grouped together (Fig. 6, regions b and d).

We discarded that sequence conservation or differences in the parsimony informative sites (*e.g.* a few nucleotide polymorphisms in a very conserved region that could theoretically lead to disparate inferences) could influence the accuracy of genotype estimation.



Fig. 2. Split decomposition and Bayesian inference results on the reconstruction of the evolutionary history of key taxa that evidence DENV-3 mosaicism. First, networks representing a 95% confidence set of topologies estimated in a Split Decomposition analysis with 100 bootstrapping replicates using (A) the complete dataset from Table 1 or (B) the representative genotyped sequences only; black numbers correspond to bootstrapping values and scale bar for distances. Second, Bayesian maximum clade credibility (MCC) trees inferred for (C) the complete dataset from Table 1 and for (*D*) the representative genotyped sequences only; black numbers correspond to posterior probability estimates, gray numbers correspond to the TMRCAs and scale bar for years.

However, we calculated the percentage of segregating sites (*S*) and nucleotide diversity values (π) (Nei and Li, 1979) in the DNASP 5.0 software (Librado and Rozas, 2009), and values were similar for all regions (*S* values of μ = 14.960, σ = 0.676 for ID-1982; μ = 15.505, σ = 1.058 for PH-1997; π values of μ = 0.05180, σ = 0.00287 for ID-1982; μ = 0.05403, σ = 0.00489 for PH-1997). Hence, we argue that a discrepancy between "genealogies" and the "genome tree" could have resulted from differential selective constraints for each region. For example, the analyses using a recombinant-free dataset often yielded a higher TMRCA estimate than the analysis using complete genomes (Fig. S5, in the supplemental material). Lastly, the phylogenetic relationships among the genotypes were uncertain with a mean value of 0.5 (σ = 0.13) for all posterior probabilities for any other sister group.

4. Discussion

4.1. DENV-3 mosaics

Recombination is one of the main forces underpinning genome diversity, bringing together diverse genomic regions with different evolutionary histories into a taxon (Awadalla, 2003). Moreover, recombination seems to be quite common in dengue and in a wide range of RNA viruses (Worobey et al., 1999). For example, Holmes et al. (1999) reported a breakpoint in the E gene of a DENV-1 taxon using strains for which contiguous C-prM-E gene sequences were available (2325 kpb). At the same time, Worobey et al. (1999) reported seven DENV strains from the four serotypes having profiles suggestive of genetic exchange strains and also advised on the apparent inadequacy of short sequence fragments to provide a complete picture of recombinant dengue virus phylogenetic relationships. By analyzing 636 genomic coding regions we found two genomes, one from Indonesia isolated in 1982 and another from the Philippines isolated in 1997, with complex mosaic recombination patterns. Although mosaicism has been previously reported for DENV (Perez-Ramirez et al., 2009), evidence for it is scant. For example, Tolou et al. (2001) and Chen et al. (2008) reported taxa with evidence for multiple recombination events using complete DENV-1 genomic sequences. In the former study, the genomic UTRs were inherited from one ancestral lineage, while an internal fragment located in the E-NS1 region originated from another. The latter study documented a recombinant taxon with



Fig. 3. The recombinant taxon ID-1982. (A) Predictions from Husmeier and Wright HMM with a four-taxa alignment; the horizontal axis shows the position in the genome, and the vertical axis represents the posterior probability for the trees shown in the left. The genomic partitions and a schematic representation of the DENV genome accompany the results positioned at the bottom. (B) Predictions from Westesson and Holmes HMM with six-taxa alignment; the grey and black backgrounds correspond to posterior probabilities of the trees shown in the left. (C) Networks representing a 95% confidence set of topologies estimated in a Split Decomposition analysis with 100 bootstrapping replicates. Most taxon names were removed for simplicity and only those of the recombinant and surrogate parents were maintained; scale bar for distances. Numbers next to region labels represent nucleotide positions.



Fig. 4. The recombinant PH-1997. (A) Predictions from Husmeier and Wright HMM with a four-taxa alignment; the horizontal axis shows the position in the genome, and the vertical axis represents the posterior probability for the trees shown in the left. The genomic partitions and a schematic representation of the DENV genome accompany the results positioned at the bottom. (B) Predictions from Westesson and Holmes HMM with six-taxa alignment; the grey and black backgrounds correspond to posterior probabilities of the trees shown in the left. (C) Networks representing a 95% confidence set of topologies estimated in a Split Decomposition analysis with 100 bootstrapping replicates. Most taxon names were removed for simplicity and only those of the recombinant and surrogate parents were maintained; scale bar for distances. Numbers next to region labels represent nucleotide positions.







Region c

Region e

0.4

59.0



ID 1988

ID 1982

24.7

0.8 24.6

30 1

7.0

GI

GIII



Region d



1.0

□ ID 1982

GII

BD 2002



Fig. 5. Maximum Clade Credibility trees for the ID-1982 dataset. Most taxa names were removed for simplicity and only those of the recombinant and parents were maintained. Black numbers correspond to posterior probability estimates and gray ones correspond to the TMRCAs; scale bar in years.

three recombination regions, located within the sequences of the prM-E junction, NS1 and NS3 genes. More recently, Carvalho et al. (2010) reported DENV-1 recombinants using complete genomic sequences, with two breakpoint positions located in the prM-E region and one within the NS5 region.

1.0

10.4

1.0

0.4 38.7

1.0

1.0

We discarded BD-2002 as a recombinant taxon because it consistently grouped within genotype II in our analyses and in previous studies (Islam et al., 2006). Interestingly, in our ML tree it occupies a basal position together with Thai isolates from 1973 and 1987 (Supplementary Fig. S1, in the supplemental material). Genome



























Fig. 6. Maximum Clade Credibility trees for the PH-1997 dataset. Most taxa names were removed for simplicity and only those of the recombinant and parents were maintained. Black numbers correspond to posterior probability estimates and gray ones correspond to the TMRCAs; scale bar in years.

Accordingly, ID-1982 and PH-1997 are recombinant taxa between: (*i*) taxa from an ancestral lineage of genotype I (related to ID-1988, MY-1997 and MY-1999) and, (*ii*) taxa from an ancestral Thai lineage of genotype II. The latter lineage could have been maintained in nature in a low circulation rate and eventually emerged in Bangladesh. Given the location of isolation of the recombinants studied here, we assume that the recombination events took place in South East Asia, where the importation and co-circulation of multiple viral lineages from all serotypes has been documented (Holmes et al., 2009; Raghwani et al., 2011; Wilder-Smith et al., 2005).

Recombination breakpoints for non-segmented [+] ssRNA viruses have been located in regions of the genome encoding for both, non-structural (Santti et al., 1999) and structural proteins (Costa-Mattioli et al., 2003). We showed that ID-1982 and PH-1997 breakpoints matched gene boundaries well, except for the NS5 gene. Recombination breakpoints, falling *between* rather than *within* genes, suggest that negative selection operated against deleterious intra-gene mosaics, as shown for HIV (Simon-Loriere et al., 2010). In the case of NS5, the recombination events appeared to abide to functional constraints, since five conserved motifs of the RdRp catalytic site fell within a unique partition (Figs. 3 and 4A and B) (Argos, 1988; Poch et al., 1989; Yap et al., 2007). The maintenance of these motifs together with their inter-motif α -helices and β -sheets (Yap et al., 2007) may be important for the correct geometric positioning and orientation of the RdRp catalytic site.

4.2. Inter genotype DENV-3 recombination

We reported recombination among sequences from different genotypes. This may generate new phenotypes by mixing virus-encoded determinants of disease severity (Uzcategui et al., 2001). Similarly, Craig et al. (2003) and Aaskov et al. (2007) reported DENV recombinants composed of portions of two different parental genotypes in a single host. However, the scarcity of data on natural recombinants suggests that such events do not generate viable strains as in the case of Hepatitis C virus (HCV) (Kalinina et al., 2002). Nevertheless, our results, together with the growing evidence of natural recombinants occurring among non segmented [+] ssRNA viruses, suggest that junction of gene regions – or functional partitions, as the catalytic sites of NS5 – may increase the likelihood of viable strains.

Our phylogenetic analyses based on distinct genomic regions further demonstrates that the existence of natural DENV recombinant strains in public databases complicates accurate genotyping. especially in the case of DENV control and molecular epidemiology. For instance, DENV genotyping by phylogenetic analysis of different genomic regions was validated in the past (Jarman et al., 2008) and in the case of DENV-3, all genes and the 3' UTR resulted in satisfactory marker genes. However, if recombinants are present, no single phylogeny describes the ancestry of nucleotide sequences under study because genomic regions will have different evolutionary histories (Degnan and Rosenberg, 2006). Besides, we have also shown that recombination further complicates the estimation of evolutionary parameters, such as the TMRCA (Awadalla, 2003). Appropriately, Posada and Crandall (2002) showed that the effect of recombination on phylogenetic inference depends on the relatedness of the sequences and on the length of the different regions involved, with recombination events being harder to detect among closely related lineages.

4.3. Mosaic structure implications for viral biology and replication

If recombination between divergent strains represents an important source of genetic variation, it may be crucial to broaden the knowledge about RNA viral replication mechanisms and, also search for other recombinants to unveil possible recurrent profiles. The detection of mosaic DENV forms implies that the DENV RdRp protein may have experienced multiple strand switching events during replication. We found an amino acid substitution at position 3293 – within the priming loop – only in the recombinants and a surrogate parental (BD-2002) and hence we hypothesized that such viruses may have an altered structure that allowed the RdRp to access and switch frequently between two strands. The priming

loop stability depends on several intra-loop interactions, and the tryptophan residue at 3293 is involved in a hydrogen bond at the center of this structure (Yap et al., 2007). The replacement of this tryptophan (nonpolar side-chain) by an arginine (positively charged side-chain) in this long and potentially flexible loop may lead to conformational changes that could facilitate strand switching. Since the priming loop regulate the access and exit of RNA templates into the catalytic site (Yap et al., 2007), we would argue that such changes may have a role in the recombination events.

Kim and Kao (2001) have shown that a recombinant RdRp from Bovine viral diarrhea virus (BVDV) can switch templates to an internal region of a different strand and then continue synthesis on the initial strand. Besides, although RNA synthesis during elongation is thought to be relatively rapid and processive (Jin et al., 2011), DENV RdRp have a higher conformational flexibility upon the transition from initiation to elongation (Selisko et al., 2006), which may favor recombination. However, this process seems to be a by-product of the processivity of RNA polymerases and not a means by which RNA viruses can purge deleterious mutations (Simon-Loriere and Holmes, 2011). Therefore, template switching could generate mainly deleterious changes and dysfunctional RNA genomes and this could help explain why just a handful of DENV recombinants have being documented so far. Nonetheless, AbuBakar et al. (2002) reported the emergence of a DENV-4 genotype IIA in Malaysia following a recombination event from different ancestral lineages; an observation that, in terms of fitness benefits, advocate for a RNA recombination process that mediates genomic arrangement leading to novel adaptive change.

The identification of recombination events concerning a human pathogen is an important discovery that requires further investigation (Tolou et al., 2001). Our evidence for DENV-3 recombinants was inferred using sequences from public databases. However, given the lack of general agreement about the significance of bioinformatics observations, and due to the lack of appropriate experimental systems, it will be appropriate to verify our observation under laboratory conditions. Experiments by Taucher et al. (2010) showed that intergenomic recombination events may occur in the structural region of flaviviruses, but its frequency appears to be very low. In addition, Chuang and Chen (2009) using site-directed mutagenesis showed that the 5' stem-loop conformation of the secondary structure is important in modulating RNA recombination of the Japanese encephalitis virus. Nonetheless, McGee et al. (2011) using flavivirus plasmid constructs highlighted that in the unlikely event of a high level acute co-infection of two distinct yellow fever genomes, the generation of viable flavivirus recombinants is extremely unlikely.

It will be also necessary to address if recombination occurs more frequently during acute human infection, in mosquitoes or both. Studies have provided evidence that DENV could persist in *Aedes* spp. during successive generations through transovarial transmission (Joshi et al., 2002; Khin and Than, 1983). Consequently, mosquito vectors could play an important role in the maintenance of DENV in nature, acting as reservoirs where concomitant infection may help spawn recombinants. Interestingly, flavivirus-related RNA sequences persist integrated in the dsDNA genome of *Aedes* spp. Mosquitoes, which implies DNA synthesis from viral RNA mediated by an endogenous reverse transcriptase activity (Crochu et al., 2004). Those findings indicate complex interactions between the mosquito's nucleic acid metabolism and DENV replication that may also contribute to the generation of recombinant viruses.

4.4. Conclusion

Our approach for exploring recombination rendered two recombinants so far. In the future, we intend to relax our screening criteria to study other plausible recombination events (*e.g.* the hits detected by four or five automated algorithms), and include the UTRs, when available. Several issues regarding DENV recombination remain unsolved, particularly concerning the diversity of recombination forms and the mechanism itself. Crucially, recombination may turn out to be a major determinant of how the virus might respond to vaccination efforts given the variety of pathogenic strains in nature (Holmes and Burch, 2000).

Our observations support the idea that natural recombinants of DENV carrying multiple recombination events (*i.e.*, DENV mosaics) do exist. We recommend further studies for testing the hypothesis that a point mutation resulted in the genetic mosaicism observed because inferences based solely on bioinformatics may be misleading in the absence of experimental validation. New results will support a model of homologous recombination, where multiple strand switching may play a role in generating genetic diversity, with implications for [+] ssRNA virus biology.

Acknowledgements

This work was supported by FAPESP (Fundação de Amparo à Pesquisa do Estado de São Paulo) grant 2010/19059-7. PMAZ holds a CNPq-PQ scholarship; AFB and CJV hold FAPESP scholarships 2011/14537-0 and 2011/17071-2, respectively.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.meegid.2013. 05.012.

References

- Aaskov, J., Buzacott, K., Field, E., Lowry, K., Berlioz-Arthaud, A., Holmes, E.C., 2007. Multiple recombinant dengue type 1 viruses in an isolate from a dengue patient, J. Gen. Virol. 88, 3334–3340.
- AbuBakar, S., Wong, P.-F., Chan, Y.-F., 2002. Emergence of dengue virus type 4 genotype IIA in Malaysia. J. Gen. Virol. 83, 2437–2442.
- Argos, P., 1988. A sequence motif in many polymerases. Nucleic Acids Res. 16, 9909–9916.
- Arnold, J.J., Cameron, C.E., 1999. Poliovirus RNA-dependent RNA polymerase (3Dpol) is sufficient for template switching in vitro. J. Biol. Chem. 274, 2706– 2716.
- Awadalla, P., 2003. The evolutionary genomics of pathogen recombination. Nat. Rev. Genet. 4, 50–60.
- Bandelt, H., Dress, A., 1992. Split decomposition: a new and useful approach to phylogenetic analysis of distance data. Mol. Phylogent. Evol., 242–252.
- Becher, P., Tautz, N., 2011. RNA recombination in pestiviruses: cellular RNA sequences in viral genomes highlight the role of host factors for viral persistence and lethal disease. RNA Biol. 8, 216–224.
- Bharaj, P., Chahar, H.S., Pandey, A., Diddi, K., Dar, L., Guleria, R., Kabra, S.K., Broor, S., 2008. Concurrent infections by all four dengue virus serotypes during an outbreak of dengue in 2006 in Delhi. India Virol. J. 5, 1.
- Bruen, T.C., Philippe, H., Bryant, D., 2006. A simple and robust statistical test for detecting the presence of recombination. Genetics 172, 2665–2681.
- Carney, J., Daly, J.M., Nisalak, A., Solomon, T., 2011. Recombination and positive selection identified in complete genome sequences of Japanese encephalitis virus. Arch. Virol. 151, 75–83.
- Carvalho, S.E.S., Martin, D.P., Oliveira, L.M., Ribeiro, B.M., Nagata, T., 2010. Comparative analysis of American dengue virus type 1 full-genome sequences. Virus Genes 40, 60–66.
- Chambers, T.J., Hahn, C.S., Galler, R., Rice, C.M., 1990. Flavivirus genome organization, expression, and replication. Annu. Rev. Microbiol. 44, 649–688.
- Chen, S.-P., Yu, M., Jiang, T., Deng, Y.-Q., Qin, C.-F., Han, J.-F., Qin, E.-D., 2008. Identification of a recombinant dengue virus type 1 with 3 recombination regions in natural populations in Guangdong province. China Arch. Virol. 153, 1175–1179.
- Cheng, C., Nagy, P., 2003. Mechanism of RNA recombination in carmo-and tombusviruses: evidence for template switching by the RNA-dependent RNA polymerase in vitro. J. Virol. 22, 12033–12047.
- Chuang, C.-K., Chen, W.-J., 2009. Experimental evidence that RNA recombination occurs in the Japanese encephalitis virus. Virology 394, 286–297.
- Cooper, P., 1974. On the nature of poliovirus genetic recombinants. J. Gen. Virol. 5, 41–49.
- Costa-Mattioli, M., Ferré, V., Casane, D., Perez-Bercoff, R., Coste-Burel, M., Imbert-Marcille, B.-M., Andre, E.C.M., Bressollette-Bodin, C., Billaudel, S., Cristina, J.,

2003. Evidence of recombination in natural populations of hepatitis A virus. Virology 311, 51–59.

- Craig, S., Thu, H.H.M., Lowry, K., Wang, X.-F., Holmes, E.C., Aaskov, J., 2003. Diverse dengue type 2 virus populations contain recombinant and both parental viruses in a single mosquito host. J. Virol. 77, 4463.
- Crochu, S., Cook, S., Attoui, H., Charrel, R.N., De Chesse, R., Belhouchet, M., Lemasson, J.-J., de Micco, P., de Lamballerie, X., 2004. Sequences of flavivirus-related RNA viruses persist in DNA form integrated in the genome of Aedes spp. mosquitoes. J. Gen. Virol. 85, 1971–1980.
- Dahourou, G., Guillot, S., Le Gall, O., Crainic, R., 2002. Genetic recombination in wildtype poliovirus. J. Gen. Virol. 83, 3103–3110.
- Degnan, J.H., Rosenberg, N.a., 2006. Discordance of species trees with their most likely gene trees. PLoS Genet. 2, e68.
- Drummond, A.J., Rambaut, A., Shapiro, B., Pybus, O.G., 2005. Bayesian coalescent inference of past population dynamics from molecular sequences. Mol. Biol. Evol. 22, 1185–1192.
- Drummond, A.J., Ho, S.Y.W., Phillips, M.J., Rambaut, A., 2006. Relaxed phylogenetics and dating with confidence. PLoS Biol. 4, e88.
- Drummond, A.J., Suchard, M.A., Xie, D., Rambaut, A., 2012. Bayesian phylogenetics with BEAUti and the BEAST 1.7. Mol. Biol. Evol. 29, 1969–1973.
- Duggal, R., Cuconati, A., Gromeier, M., Wimmer, E., 1997. Genetic recombination of poliovirus in a cell-free system. Proc. Natl. Acad. Sci. USA 94, 13786–13791. Edgar, R.C., 2004a. MUSCLE: a multiple sequence alignment method with reduced
- time and space complexity. BMC Bioinformatics 5, 113.
- Edgar, R.C., 2004b. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32, 1792–1797.
- Felsenstein, J., Churchill, G.A., 1995. A hidden Markov model approach evolution to variation among sites in rate of evolution. Mol. Biol. Evol. 13, 93–104.
- Friedman, N., Ninio, M., Pe'er, I., Pupko, T., 2002. A structural EM algorithm for phylogenetic inference. J. Comput. Biol. 9, 331–353.
- Gmyl, A.P., Belousov, E.V., Maslova, S.V., Khitrina, E.V., Chetverin, A.B., Agol, V.I., 1999. Nonreplicative RNA recombination in Poliovirus. J. Virol. 73, 8958–8965.
- Gould, E.A., Gritsun, T.S., 2006. Taxonomy and evolution of the flaviviridae. In: Kalitzky, M., Borowski, P. (Eds.), Molecular Biology of the Flavivirus. Horizon Bioscience, Norfolk, UK.
- Gould, E.A., Moss, S.R., Turner, S.L., 2004. Evolution and dispersal of encephalitic flaviviruses. Arch. Virol. Suppl. 18, 65–84.
- Holmes, E.C., Burch, S.S., 2000. The causes and consequences of genetic variation in dengue virus. Trends Microbiol. 8, 74–77.
- Holmes, E.C., Worobey, M., Rambaut, A., 1999. Phylogenetic evidence for recombination in dengue virus. Mol. Biol. Evol. 16, 405–409.
- Holmes, E.C., Tio, P.H., Perera, D., Muhi, J., Cardosa, J., 2009. Importation and cocirculation of multiple serotypes of dengue virus in Sarawak. Malaysia Virus Res. 143, 1–5.
- Husmeier, D., McGuire, G., 2003. Detecting recombination in 4-taxa DNA sequence alignments with bayesian hidden Markov models and Markov chain Monte Carlo. Mol. Biol. Evol. 20, 315–337.
- Huson, D.H., Kloepper, T.H., 2005. Computing recombination networks from binary sequences. Bioinformatics 21 (Suppl. 2), 159–165.
- Islam, M.A., Ahmed, M.U., Begum, N., Chowdhury, N.A., Khan, A.H., Parquet, M.D.C., Bipolo, S., Inoue, S., Hasebe, F., Suzuki, Y., Morita, K., 2006. Molecular characterization and clinical evaluation of dengue outbreak in 2002 in Bangladesh. Jpn. J. Infect. Dis. 59, 85–91.
- Jarman, R.G.R., Holmes, E.E.C., Rodpradit, P., Klungthong, C., Gibbons, R.V., Nisalak, A., Rothman, A.L., Libraty, D.H., Ennis, F.A., Mammen, M.P., Endy, T.P., 2008. Microevolution of Dengue viruses circulating among primary school children in kamphaeng phet. Thailand J. Virol. 82, 1–23.
- Jin, Z., Deval, J., Johnson, K.a., Swinney, D.C., 2011. Characterization of the elongation complex of dengue virus RNA polymerase: assembly, kinetics of nucleotide incorporation, and fidelity. J. Biol. Chem. 286, 2067–2077.
- Joshi, V., Mourya, D.T., Sharma, R.C., 2002. Persistence of dengue-3 virus through transovarial transmission passage in successive generations of Aedes aegypti mosquitoes. Am. J. Trop. Med. Hyg. 67, 158–161.
- Kalinina, O., Norder, H., Mukomolov, S., Magnius, L.O., 2002. A natural intergenotypic recombinant of hepatitis C virus identified in St. Petersburg J. Virol. 76, 4034–4043.
- Karabatsos, N., 1985. International catalogue of arthropod-borne viruses. A.S.f.T.M.a. Hygiene, San Antonio, TX.
- Khin, M., Than, K., 1983. Transovarial transmission of dengue 2 virus by Aedes aegypti in nature. Am. J. Trop. Med. Hyg. 32, 590–594.
- Kim, M.J., Kao, C., 2001. Factors regulating template switch in vitro by viral RNAdependent RNA polymerases: implications for RNA-RNA recombination. Proc. Natl. Acad. Sci. USA 98, 4972–4977.
- Laille, M., Deubel, V., Sainte-Marie, F.F., 1991. Demonstration of concurrent dengue 1 and dengue 3 infection in six patients by the polymerase chain reaction. J. Med. Virol. 34, 51–54.
- Lavergne, A., de Thoisy, B., Lacoste, V., Pascalis, H., Pouliquen, J.-F., Mercier, V., Tolou, H., Dussart, P., Morvan, J., Talarmin, A., Kazanji, M., 2006. Mayaro virus: complete nucleotide sequence and phylogenetic relationships with other alphaviruses. Virus Res. 117, 283–290.
- Librado, P., Rozas, J., 2009. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics 25, 1451–1452.
- Lindenbach, B.D., Rice, C.M., 2003. Molecular biology of flaviviruses. Adv. Virus Res. 59, 23–61.
- Loroño-Pino, M.A., Cropp, C.B., Farfán, J.A., Vorndam, A.V., Rodríguez-Angulo, E.M., Rosado-Paredes, E.P., Flores-Flores, L.F., Beaty, B.J., Gubler, D.J., 1999. Common

occurrence of concurrent infections by multiple dengue virus serotypes. Am. J. Trop. Med. Hyg. 61, 725–730.

- Martin, D.P., Lemey, P., Lott, M., Moulton, V., Posada, D., Lefeuvre, P., 2010. RDP3: a flexible and fast computer program for analyzing recombination. Bioinformatics 26, 2462–2463.
- McGee, C.E., Tsetsarkin, K.A., Guy, B., Lang, J., Plante, K., Vanlandingham, D.L., Higgs, S., 2011. Stability of yellow fever virus under recombinatory pressure as compared with chikungunya virus. PLoS One 6, e23247.
- Milne, I., Lindner, D., Bayer, M., Husmeier, D., McGuire, G., Marshall, D.F., Wright, F., 2009. TOPALi v2: a rich graphical interface for evolutionary analyses of multiple alignments on HPC clusters and multi-core desktops. Bioinformatics 25, 126– 127.
- Montenegro, F., Araújo, D.C., Maria, R., Nogueira, R., Maria, J., Araújo, G.D., Letícia, I., Ramalho, C., Lucia, M., Sá, F.D., Elizabeth, M., Melo, L.D., Castelo, I., Coelho, B., 2006. Concurrent infection with dengue virus type-2 and DENV-3 in a patient from Ceará. Brazil Mem. Inst. Oswaldo Cruz 101, 925–928.
- Nagy, P., Simon, A., 1997. New insights into the mechanisms of RNA recombination. Virology 9, 1–9.
- Nájera, R., Delgado, E., Pérez-Alvarez, L., Thomson, M.M., 2002. Genetic recombination and its role in the development of the HIV-1 pandemic. AIDS 16 (Suppl. 4), S3–S16.
- Nei, M., Li, W.H., 1979. Mathematical model for studying genetic variation in terms of restriction endonucleases. Proc. Natl. Acad. Sci. USA 76, 5269–5273.
- Perez-Ramirez, G., Diaz-Badillo, A., Camacho-Nuez, M., Cisneros, A., Munoz, M.D.L., 2009. Multiple recombinants in two dengue virus, serotype-2 isolates from patients from Oaxaca. Mexico BMC Microbiol. 9, 260.
- Poch, O., Sauvaget, I., Delarue, M., Tordo, N., 1989. Identification of four conserved motifs among the RNA-dependent polymerase encoding elements. EMBO J. 8, 3867–3874.
- Posada, D., Crandall, K.a., 2002. The effect of recombination on the accuracy of phylogeny estimation. J. Mol. Evol. 54, 396–402.
- Raghwani, J., Rambaut, A., Holmes, E.C., Hang, V.T., Hien, T.T., Farrar, J., Wills, B., Lennon, N.J., Birren, B.W., Henn, M.R., Simmons, C.P., 2011. Endemic dengue associated with the co-circulation of multiple viral lineages and localized density-dependent transmission. PLoS Pathog. 7, e1002064.
- Reiter, J., Pérez-Vilaró, G., Scheller, N., Mina, L.B., Díez, J., Meyerhans, A., 2011. Hepatitis C virus RNA recombination in cell culture. J. Hepatol. 55, 777–783.
- Rice, C., Lenches, E., Shin, S., Sheets, R., Strauss, J., 1985. Nucleotide sequence of yellow fever virus: implications for flavivirus gene expression and evolution. Science 229.
- Santti, J., Hyypiä, T., Kinnunen, L., 1999. Evidence of recombination among enteroviruses. J. Virol. 73, 8741–8749.
- Selisko, B., Dutartre, H., Guillemot, J.-C., Debarnot, C., Benarroch, D., Khromykh, A., Desprès, P., Egloff, M.-P., Canard, B., 2006. Comparative mechanistic studies of de novo RNA synthesis by flavivirus RNA-dependent RNA polymerases. Virology 351, 145–158.

- Sibold, C., Meisel, H., Krüger, D.H., Labuda, M., Lysy, J., Kozuch, O., Pejcoch, M., Vaheri, A., Plyusnin, A., 1999. Recombination in Tula hantavirus evolution: analysis of genetic lineages from Slovakia. J. Virol. 73, 667–675.
- Simon, A.E., Nagy, P.D., 1996. RNA recombination in turnip crinkle virus: its role in formation of chimeric RNAs, multimers, and in 3 ' -end repair. Semin. Virol. 7, 373–379.
- Simon-Loriere, E., Holmes, E.C., 2011. Why do RNA viruses recombine? Nat. Rev. Microbiol. 9, 617–626.
- Simon-Loriere, E., Martin, D.P., Weeks, K.M., Negroni, M., 2010. RNA structures facilitate recombination-mediated gene swapping in HIV-1. J. Virol. 84, 12675– 12682.
- Taucher, C., Berger, A., Mandl, CW, 2010. A trans-complementing recombination trap demonstrates a low propensity of flaviviruses for intermolecular recombination. J virol. 84, 599–611.
- Thai, K.T.D., Henn, M.R., Zody, M.C., Tricou, V., Nguyet, N.M., Charlebois, P., Lennon, N.J., Green, L., de Vries, P.J., Hien, T.T., Farrar, J., van Doorn, H.R., de Jong, M.D., Birren, B.W., Holmes, E.C., Simmons, C.P., 2012. High-resolution analysis of intrahost genetic diversity in dengue virus serotype 1 infection identifies mixed infections. J. Virol. 86, 835–843.
- Tolou, H.J., Couissinier-Paris, P., Durand, J.P., Mercier, V., de Pina, J.J., de Micco, P., Billoir, F., Charrel, R.N., de Lamballerie, X., 2001. Evidence for recombination in natural populations of dengue virus type 1 based on the analysis of complete genome sequences. J. Gen. Virol. 82, 1283–1290.
- Uzcategui, N.Y., Camacho, D., Comach, G., Cuello de Uzcategui, R., Holmes, E.C., Gould, E.A., 2001. Molecular epidemiology of dengue type 2 virus in Venezuela: evidence for in situ virus evolution and recombination. J. Gen. Virol. 82, 2945– 2953.
- Wang, W.-K., Chao, D.-Y., Lin, S.-R., King, C.-C., Chang, S.-C., 2003. Concurrent infections by two dengue virus serotypes among dengue patients in Taiwan. J. Microbiol. Immunol. Infect. 36, 89–95.
- Weaver, S.C., Vasilakis, N., 2009. Molecular evolution of dengue viruses: contributions of phylogenetics to understanding the history and epidemiology of the preeminent arboviral disease. Infect. Genet. Evol. 9, 523–540.
- Wilder-Smith, A., Yoksan, S., Earnest, A., Subramaniam, R., Paton, N.I., 2005. Serological evidence for the co-circulation of multiple dengue virus serotypes in Singapore. Epidemiol. Infect. 133, 667–671.
- Worobey, M., Rambaut, a., Holmes, E.C., 1999. Widespread intra-serotype recombination in natural populations of dengue virus. Proc. Natl. Acad. Sci. USA 96, 7352–7357.
- Yap, T.L., Xu, T., Chen, Y.-L., Malet, H., Egloff, M.-P., Canard, B., Vasudevan, S.G., Lescar, J., 2007. Crystal structure of the dengue virus RNA-dependent RNA polymerase catalytic domain at 1.85-angstrom resolution. J. Virol. 81, 4753– 4765.
- Zwickl, D.J., 2006. Genetic Algorithm Approaches for the Phylogenetic Analysis of Large Biological Sequence Datasets Under the Maximum Likelihood Criterion. The University of Texas, Austin.