The Pangenome of the *Anticarsia gemmatalis* Multiple Nucleopolyhedrovirus (AgMNPV)

Anderson Fernandes de Brito1, Carla Torres Braconi1, Manfred Weidmann2, Meik Dilcher2, João Marcelo Pereira Alves3, Arthur Grüber3, and Paolo Marinho de Andrade Zanotto1,*

1Department of Microbiology, Institute of Biomedical Sciences—ICB II, Laboratory of Molecular Evolution and Bioinformatics, University of São Paulo—USP, São Paulo, SP, Brazil
2Department of Virology, University Medical Center, Göttingen, Germany
3Department of Parasitology, Institute of Biomedical Sciences—ICB II, University of São Paulo—USP, São Paulo, SP, Brazil
*Corresponding author: E-mail: pzanotto@usp.br.

Data deposition: This project has been deposited at GenBank under 17 accession numbers, from KR815455 to KR815471.

Accepted: November 16, 2015

Abstract

The alphabaculovirus *Anticarsia gemmatalis* multiple nucleopolyhedrovirus (AgMNPV) is the world’s most successful viral bioinsecticide. Through the 1980s and 1990s, this virus was extensively used for biological control of populations of *Anticarsia gemmatalis* (Velvetbean caterpillar) in soybean crops. During this period, genetic studies identified several variable loci in the AgMNPV; however, most of them were not characterized at the sequence level. In this study we report a full genome comparison among 17 wild-type isolates of AgMNPV. We found the pangenome of this virus to contain at least 167 hypothetical genes, 151 of which are shared by all genomes. The gene *bro-a* that might be involved in host specificity and carrying transporter is absent in some genomes, and new hypothetical genes were observed. Among these genes there is a unique *mf12-like* gene, probably implicated in ubiquitination. Events of gene fission and fusion are common, as four genes have been observed as single or split open reading frames. Gains and losses of genomic fragments (from 20 to 900 bp) are observed within tandem repeats, such as in eight direct repeats and four homologous regions. Most AgMNPV genes present low nucleotide diversity, and variable genes are mainly located in a locus known to evolve through homologous recombination. The evolution of AgMNPV is mainly driven by small indels, substitutions, gain and loss of nucleotide stretches or entire coding sequences. These variations may cause relevant phenotypic alterations, which probably affect the infectivity of AgMNPV. This work provides novel information on genomic evolution of the AgMNPV in particular and of baculoviruses in general.

Key words: baculovirus, evolution, wild isolates, horizontal gene transfer, deep sequencing, genetic diversity.

Introduction

Baculoviruses are insect-specific viruses that are frequently restricted to one or a few related insect species, and thus became widely applied as biological controls against pests in agriculture and forestry (O’Reilly et al. 1993; Tanada and Kaya 1993; Moscardi 1999; Szewczyk et al. 2006). Baculoviruses have a circular, covalently closed, double-stranded DNA genome and, up to now, the most studied genera are the Alpha and Betabaculovirus. The Alpha-baculovirus *Anticarsia gemmatalis* multiple nucleopolyhedrovirus (AgMNPV) was discovered in Peru, in 1962 (Steinhaus and Marsh 1962), and its first occurrence in Brazilian territory was registered in 1972 (Allen and Knell 1977). In Brazil, during the 1980s and 1990s, AgMNPV was extensively applied in soybean crops against populations of velvetbean caterpillars (*A. gemmatalis*), and these actions are recognized as the world’s largest biological control program using a virus as a bioinsecticide (Moscardi 1989, 1999). In nature, multiple viral particles of this virus are found embedded in paracrystalline protein structures called occlusion bodies (OBs) (Rohrmann 2011). When a caterpillar ingests these structures, the alkaline environment of the larva midgut promotes OBs dissolution and occlusion-derived viruses (ODVs) are released, establishing primary infection. Within the first 2 hours the first budded viruses (BVs) are produced, and secondary infection takes place systemically, giving rise to new OBs (Friesen and Miller 2001). In most
baculoviruses the infection cycle ends with insect disintegration (melting), a process that leads to the release of OBs into the environment. Interestingly, AgMNPV lacks the genes responsible for host cuticle rupture and tissue liquefaction (cathepsin and chitinase), and this fact favors the collection of dead insects, useful for future biological control application (Slack et al. 2004). In 2006, the isolate 2D was the first plaque-purified isolate of AgMNPV to have its whole genome sequenced (accession number DQ813662) (Oliveira et al. 2006). It contains 152 open-reading frames (ORFs), with short or absent intergenic regions, nine homologous regions (hrs), and shows high similarity to CDefMNPV (Oliveira et al. 2006). Previous studies on AgMNPV diversity documented high levels of polymorphism in some genomic loci; however, the specific genetic elements affected by those changes, as well as their impact on AgMNPV evolution, are mostly unknown (Maruniak 1989; Maruniak et al. 1999). Wild populations of baculoviruses are commonly composed of different genotypes, which frequently recombine in a context of coinfection (Croizier and Ribeiro 1992). The genome evolution of baculoviruses involves gene horizontal gene transfer (HGT) (Hughes and Friedman 2003); insertion of noncoding sequences of cellular origin (transposons) (Blissard and Rohrmann 1990); and duplication/loss of genes or genomic fragments (Hayakawa et al. 2000), following a processes by which genomes tend to undergo net gain and losses of accessory genes, showing deviation from linkage disequilibrium among functional categories (Zanotto and Krakauer 2008). In this study, we describe the complete genome sequence and analyses of 17 naturally occurring South American isolates of AgMNPV. We report the main events outlining the AgMNPV genomic evolution, such as gene gain/loss, gene fusion/fission, among other structural variations. These analyses provide a better understanding of AgMNPV evolution and their molecular biology and encourage the search for new baculoviruses potentially able to be used in biological pest control.

**Materials and Methods**

**Viral Samples, In Vivo Amplification, and DNA Extraction**

The 17 wild-type samples were isolated from *A. gemmatalis* larvae. These insects had been collected from soybean crops with no previous intentional usage of baculoviruses as biological control agents. These samples were kindly provided by Dr Flavio Moscardi and Daniel Sosa-Gómez (EMBRAPA Soja, Londrina, Brazil). These isolates are wild-type genetically heterogeneous viral populations of AgMNPV, obtained from different geographical locations from 1980 to 1990, in southern Brazil, Argentina, and Uruguay (fig. 1 and table 1). The viral stocks were amplified separately in vivo (per os) in third/fourth-instar *A. gemmatalis* caterpillars fed on an artificial diet, as previously described (Greene et al. 1976; O’Reilly et al. 1993). The propagated polyhedra were purified from insect cadavers by steps of maceration, centrifugation and sucrose gradient ultracentrifugation, as previously described (Maruniak 1986). After purification, polyhedra were treated with DNase, to prevent any contamination with insect host genome. Virus particles were solubilized under alkaline conditions (Whitt and Manning 1987); genomic DNA was purified by successive phenol/chloroform/isoamyl alcohol extractions followed by ethanol precipitation and elution in TE buffer (Lo et al. 1996).

**454 Next-Generation Sequencing**

First, to prepare viral DNA libraries, approximately 50 ng of genomic DNA from each sample were separately fragmented using the Roche Titanium-compatible Nextera DNA Sample Prep Kit (Epicentre Biotechnologies), according to the manufacturer’s protocol. In different reactions, viral DNA were MID-tagged (Roche 454 Life Sciences) during the Nextera amplification steps. The libraries were purified with Agencourt AMPure XP Beads (Beckman Coulter, Inc), at a 1:0.7 ratio (DNA:Beads) to eliminate fragments below 300 bp. The purified material was quantified and submitted to emPCR reactions, following the manufacturer’s instructions (emPCR Method Manual—Lib-L SV, Roche 454 Life Sciences). Finally, those pools were sequenced using a GS FLX Titanium platform (Roche 454 Life Sciences), in a single run, in accordance with manufacturer’s recommendations (Sequencing Method Manual, GS FLX Titanium Series, Roche 454 Life Sciences). The specific reads of each sample were segregated into 17

---

**Fig. 1.**—Geographic regions of original field sampling of the AgMNPV isolates (approximate coordinates).
SFF files according to their MID tags, which were removed at this step.

Genome Assembly

There is considerable variability in wild-type, field isolates of natural organisms that did not experience significant, recent bottlenecks. Nevertheless, to facilitate comparisons, we aimed at the reconstruction of the predominant genotype of each viral population. This was done by de novo genome assemblies using the “Geneious” proprietary assembler implemented in Geneious 7.1.7 (Kearse et al. 2012). Summarizing, sequencing reads had their adapters and low quality regions trimmed before the assembly. The genomes were then assembled taking into account two parameters: Minimum overlap of 150 nt and 97% of identity among reads. The most frequent haplotypes reconstructed in large contigs were joined to obtain 17 full-length genomic scaffolds. 454-sequencing errors at homopolymeric regions were manually inspected and individually corrected.

Information about total number of reads, average read length, and coverage can be found on supplementary figure S1, Supplementary Material online.

Genome Annotation and Sequence Data Analysis

Each genome was annotated using a pipeline generated by the EGene 2 platform (Durham et al. 2005). In this pipeline, coding regions were predicted ab initio using Glimmer 3.02 (Delcher et al. 1999), using all coding sequences of alphabaculoviruses Group I as a training set. All ORFs with at least 138 bp and presenting maximum overlapping of 148 bp between adjacent ORFs were annotated. All translated products were submitted to similarity searches with BLASTp against NCBI’s nr (nonredundant) database. Hits were considered positive when presenting e values below $10^{-6}$. Protein domains and families were identified through searches on InterPro (Mitchell et al. 2014) and CDD (Marchler-Bauer et al. 2007) databases. Gene Ontology (GO) terms of the three ontologies were assigned to each coding sequence (Birnrs et al. 2009). All translated products were also classified into orthologous groups according to eggNOG (Muller et al. 2010) and KEGG Orthology (KO) (Aoki-Kinoshita and Kanehisa 2007), and mapped onto KEGG pathways. Additional analyses included TMHMM 2.0 (Krogh et al. 2001), Phobius1.01 (Kall et al. 2004), and SignalP4.0 (Petersen et al. 2011) to predict signal peptides, protein motifs, and secondary structures (e.g., transmembrane helices, etc.). Finally, we also performed manual curation to include missing ORFs.

hrs were detected using Dotter (Sonnhammer and Durbin 1995), whereas Tandem Repeats Finder (Benson 1999) was applied to accurately locate hrs and direct repeats (drs) in the genomes. To analyze the genetic diversity present in the populations of geographical isolates of AgMNPV, we did alignments of gene coding regions with ClustalW (Thompson et al. 1994), which were further edited manually with Bioedit (Hall 1999). The genetic diversity was measured with DnaSP 5.0 (Librado and Rozas 2009) using the curated alignments. The DnaSP 5.0 software implements the Watterson’s estimator ($\theta = Ne \mu$, assuming baculoviruses to be haploid) taking into account the number of segregating sites per gene (Watterson 1975). Moreover, a whole-genome alignment of all 17 AgMNPV genomes was performed using the

<table>
<thead>
<tr>
<th>Isolate</th>
<th>Accession Number</th>
<th>Country (Municipality, State)</th>
<th>Year</th>
</tr>
</thead>
<tbody>
<tr>
<td>AgMNPV-26</td>
<td>KR815455</td>
<td>Brazil (Dourados, Mato Grosso do Sul)</td>
<td>1984</td>
</tr>
<tr>
<td>AgMNPV-27</td>
<td>KR815456</td>
<td>Argentina</td>
<td>1982</td>
</tr>
<tr>
<td>AgMNPV-28</td>
<td>KR815457</td>
<td>Uruguay</td>
<td>1984</td>
</tr>
<tr>
<td>AgMNPV-29</td>
<td>KR815458</td>
<td>Brazil (Ponta Grossa, Paraná)</td>
<td>1985</td>
</tr>
<tr>
<td>AgMNPV-30</td>
<td>KR815459</td>
<td>Uruguay</td>
<td>1982</td>
</tr>
<tr>
<td>AgMNPV-31</td>
<td>KR815460</td>
<td>Brazil (Sertanópolis, Paraná)</td>
<td>1986</td>
</tr>
<tr>
<td>AgMNPV-32</td>
<td>KR815461</td>
<td>Brazil (Dourados, Mato Grosso do Sul)</td>
<td>1982</td>
</tr>
<tr>
<td>AgMNPV-33</td>
<td>KR815462</td>
<td>Brazil (Pelotas, Rio Grande do Sul)</td>
<td>1984</td>
</tr>
<tr>
<td>AgMNPV-34</td>
<td>KR815463</td>
<td>Brazil (Londrina, Paraná)</td>
<td>1990</td>
</tr>
<tr>
<td>AgMNPV-35</td>
<td>KR815464</td>
<td>Brazil (Rio Grande do Sul)</td>
<td>1984</td>
</tr>
<tr>
<td>AgMNPV-36</td>
<td>KR815465</td>
<td>Brazil (Passo Fundo, Rio Grande do Sul)</td>
<td>1982</td>
</tr>
<tr>
<td>AgMNPV-37</td>
<td>KR815466</td>
<td>Brazil (Campos Mourao, Paraná)</td>
<td>1982</td>
</tr>
<tr>
<td>AgMNPV-38</td>
<td>KR815467</td>
<td>Uruguay</td>
<td>1982</td>
</tr>
<tr>
<td>AgMNPV-39</td>
<td>KR815468</td>
<td>Brazil (Sertanópolis, Paraná)</td>
<td>1987</td>
</tr>
<tr>
<td>AgMNPV-40</td>
<td>KR815469</td>
<td>Brazil (Londrina, Paraná)</td>
<td>N/Aa</td>
</tr>
<tr>
<td>AgMNPV-42</td>
<td>KR815470</td>
<td>Brazil (Londrina, Paraná)</td>
<td>1987</td>
</tr>
<tr>
<td>AgMNPV-43</td>
<td>KR815471</td>
<td>Brazil (Londrina, Paraná)</td>
<td>1980</td>
</tr>
</tbody>
</table>

Note.—Isolate, location, year of collection and accession number are listed in the table.

aYear of isolation unknown.
bGeographic location of origin not specified.
“mauveAligner” algorithm implanted in Geneious 7.1.7 (Kearse et al. 2012), which in turn was applied to detect nonsynonymous changes. Both results of genetic diversity were plotted in a circular map using Circos (Krzywinski et al. 2009). Gene splicing in baculovirus appears to be rare and was reported in the ie-01ie-1 genes of the AcMNPV (Chisholm and Henner et al. 1988; Chen et al. 2013). Nevertheless, similarly to what has been done for most baculovirus-annotated genomes available in GenBank to this date, we chose to not annotate splicing because RNA-Seq data were not available for the precise identification of intron boundaries, and the lack of sufficient comparative data would hamper the construction of reliable training sets for gene splicing prediction.

New Genes in the Genomes of the AgMNPV

We coupled our previous proteomics results and findings with new sequences obtained from the AgMNPV isolates. Briefly, to conduct our analysis we used the MASCOT 2.0 online software (MatrixScience, Boston, MA) and the raw data files generated during the proteome study of AgMNPV (Braconi et al. 2014). We searched these data against the database containing the sequences of 17 wild-type genomes of AgMNPV. All parameters in the MASCOT software were set according to the sequences of 17 wild-type genomes of AgMNPV. All AgMNPV isolates reported in the supplementary fig. S1, Supplementary Material online). Following assembly, the size of the double-stranded circular DNA of AgMNPV genome varied, and we determined the similarities/differences at sequence level between all isolates. We also considered the fact that Roche/454 platform can induce erroneous base calls on homopolymeric regions, so all single nucleotide polymorphisms that mapped to those regions were ignored.

Results

Deep Sequencing of AgMNPV Genomes

The 17 AgMNPV viral populations were isolated from dead A. gemmatalis caterpillars and their genomes were sequenced using the next-generation technology Roche 454 GS-FLX Titanium. Each data set contained 26,474–78,766 reads, with an average length varying from 326 to 369 bp (supplementary fig. S1, Supplementary Material online). Following assembly, the size of the double-stranded circular DNA of AgMNPV genome varied, and we determined the similarities/differences at sequence level between all isolates. We also considered the fact that Roche/454 platform can induce erroneous base calls on homopolymeric regions, so all single nucleotide polymorphisms that mapped to those regions were ignored.

General Aspects of the AgMNPV Genomes

Compared with the AgMNPV-2D genome used as reference, the 17 wild-type genomes are from 0.04% to 1.31% smaller in length, their GC content varies from 44.5% to 44.6%, and their sequence similarity against the reference genome ranged from 98.9% to 99.5% (table 2). All AgMNPV isolates remained collinear to each other across their evolutionary history. No rearrangements, such as inversions and translocations, were detected, but insertions and deletions were common throughout the genomes (see below for more details). Most intergenic spacers in AgMNPV have not shown large modifications. The reference isolate (2D) has a highly compact genome, where 101 ORFs are separated from their neighboring ORFs by less than 50 nt, and 67 ORFs share from one to more than 140 nt with adjacent ORFs. This pattern of genome compaction is conserved in most isolates, and mutations in noncoding regions were mainly observed within loci of drs, hrs, and on regions involved in gain or loss of genes.

The aforementioned genomic features of AgMNPV were taken into account for genome annotation: Only ORFs of at least 138 bp in length (as observed in AgMNPV-2D, ORF010) and maximum overlap of 148 bp with adjacent ORFs (as in the anti-apoptotic 2, iap-2, ORF092) were annotated in the 17 AgMNPV wild-type isolates. Considering these parameters, a total of 167 ORFs were found in the pangenome of AgMNPV (fig. 2 and table 2). Among the 152 ORFs previously assigned to the AgMNPV-2D genome, 145 are shared by all isolates, but eight of them have size variations: 43 132,077 44.6 159 11 10 99.5 2D 132,239 44.5 158 10 12 — 26 131,678 44.6 157 10 12 99.5 27 131,172 44.6 157 11 12 98.9 28 130,745 44.6 157 11 12 99.2 29 130,506 44.6 157 11 12 98.9 30 130,741 44.5 156 10 13 99.3 31 132,126 44.6 158 11 12 99.2 32 131,494 44.6 157 10 12 99.4 33 131,059 44.5 157 10 12 99.4 34 131,543 44.6 158 11 10 99.1 35 132,176 44.6 159 11 11 99.3 36 131,216 44.5 156 10 12 99.3 37 131,855 44.5 156 10 13 99.3 38 130,740 44.5 156 10 13 99.3 39 130,699 44.5 157 10 12 99.3 40 132,180 44.6 158 11 12 99.3 41 130,949 44.6 157 11 13 99.2 42 132,077 44.6 159 11 10 99.5

Table 2

<table>
<thead>
<tr>
<th>Isolate</th>
<th>Size (bp)</th>
<th>GC Content (%)</th>
<th>Number of ORFs</th>
<th>Number of hrs</th>
<th>Number of drs</th>
<th>Similarity with AgMNPV-2D (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2D</td>
<td>132,239</td>
<td>44.5</td>
<td>158</td>
<td>10</td>
<td>12</td>
<td>—</td>
</tr>
<tr>
<td>26</td>
<td>131,678</td>
<td>44.6</td>
<td>157</td>
<td>10</td>
<td>12</td>
<td>99.5</td>
</tr>
<tr>
<td>27</td>
<td>131,172</td>
<td>44.6</td>
<td>157</td>
<td>11</td>
<td>12</td>
<td>98.9</td>
</tr>
<tr>
<td>28</td>
<td>130,745</td>
<td>44.6</td>
<td>157</td>
<td>11</td>
<td>12</td>
<td>99.2</td>
</tr>
<tr>
<td>29</td>
<td>130,506</td>
<td>44.6</td>
<td>157</td>
<td>11</td>
<td>12</td>
<td>98.9</td>
</tr>
<tr>
<td>30</td>
<td>130,741</td>
<td>44.5</td>
<td>156</td>
<td>10</td>
<td>13</td>
<td>99.3</td>
</tr>
<tr>
<td>31</td>
<td>132,126</td>
<td>44.6</td>
<td>158</td>
<td>11</td>
<td>12</td>
<td>99.2</td>
</tr>
<tr>
<td>32</td>
<td>131,494</td>
<td>44.6</td>
<td>157</td>
<td>10</td>
<td>12</td>
<td>99.4</td>
</tr>
<tr>
<td>33</td>
<td>131,059</td>
<td>44.5</td>
<td>157</td>
<td>10</td>
<td>12</td>
<td>99.4</td>
</tr>
<tr>
<td>34</td>
<td>131,543</td>
<td>44.6</td>
<td>158</td>
<td>11</td>
<td>10</td>
<td>99.1</td>
</tr>
<tr>
<td>35</td>
<td>132,176</td>
<td>44.6</td>
<td>159</td>
<td>11</td>
<td>11</td>
<td>99.3</td>
</tr>
<tr>
<td>36</td>
<td>131,216</td>
<td>44.5</td>
<td>156</td>
<td>10</td>
<td>12</td>
<td>99.3</td>
</tr>
<tr>
<td>37</td>
<td>131,855</td>
<td>44.5</td>
<td>156</td>
<td>10</td>
<td>13</td>
<td>99.3</td>
</tr>
<tr>
<td>38</td>
<td>130,740</td>
<td>44.5</td>
<td>156</td>
<td>10</td>
<td>13</td>
<td>99.3</td>
</tr>
<tr>
<td>39</td>
<td>130,699</td>
<td>44.5</td>
<td>157</td>
<td>10</td>
<td>12</td>
<td>99.3</td>
</tr>
<tr>
<td>40</td>
<td>132,180</td>
<td>44.6</td>
<td>158</td>
<td>11</td>
<td>12</td>
<td>99.3</td>
</tr>
</tbody>
</table>

Table 2 Genomic Features of AgMNPV Genotypes

In the end, the validated proteins had at least two independent spectra with a minimal length of ten amino acids, with greater than 99.0% probability estimated by the Peptide Prophet algorithm (Kelber et al. 2002).
codes for a structural protein at nucleocapsid base and odv-e66 codes for an ODV envelope protein. Due to structural polymorphisms leading to ORF fission, fusion or indels, some ORFs typically observed in AgMNPV-2D are not present in all genomes, namely pe-38-like-1 (ORF006), pe-38-like-2 (ORF007), he65-like-1 (ORF060), he65-like-2 (ORF061), ag144 (ORF144), bro-c (ORF156), and bro-a (ORF163). Completing the set of 167 ORFs, six additional ORFs present but not previously annotated in AgMNPV were found in all genomes (ORFs 027, 035, 042, 058, 112, and 135); and three new ORFs derived from indel events were observed in some isolates (ORFs 008, 012, and 139). To sum up, AgMNPV isolates have 151 genes in common, and besides coding sequences, a nonredundant set of 14 drs and 11 hrs were identified among all AgMNPV genomes (fig. 2 and table 2). For more details on presence or absence of ORFs, drs, and hrs, see supplementary tables S1 and S2, Supplementary Material online.

Structural Variations in hrs and drs

In the pangenome of AgMNPV, most of the DNA length variations (indels) were observed within tandem repeats, such as hrs and drs. The 14 drs consist of two or more repeating units ranging in length from 14 to 28 bp. They are minisatellites found in both intra- and intergenic regions, and most indels in coding sequence were found within these regions, especially in 1629capsid (ORF002, dr1 and dr2), pe-38 (ORF005, dr3), and ac91-like (ORF074, dr6 and dr7), where the polymorphisms gave rise to different alleles varying in hundreds of base pairs. Only four drs were found in intergenic regions, being dr4 the longer and more variable of these elements, ranging from 344 to 525 bp.

A total of 11 hrs were identified in AgMNPV, and nine of them correspond to those previously found in AgMNPV-2D. Two of them correspond to new hrs composed of singleton palindromes (hr1 and hr2), which were observed in similar
genomic positions as hr13 and hr12 in the CDEFMNPV (Lauzon et al. 2005). Four hrs (hr3, hr9, hr10, and hr11) had size variation related to gain or loss of 28-bp core palindromes. Remarkable variations were observed in hr3 (former hr4 in AgMNPV-2D), which had four variants, ranging from 342 (5 × 30 bp imperfect palindromes) to 1,232 bp (19 × 30 bp imperfect palindromes) (supplementary table S2, Supplementary Material online).

Structural Variations in Coding Sequences

Some AgMNPV genes underwent insertions and deletions comprising stretches varying from dozens to hundreds of nucleotides, which caused structural variations, such as gene shortening and truncation. A large 295-bp indel generated a truncated form of the hsp (ORF136), a variation unique to the AgMNPV. This truncated allele potentially encodes a disrupted HAD-like domain, which shows a putative histidinol-phosphatase activity (Cohen et al. 2009), encoded by at least eight isolates: 26, 28, 29, 30, 36, 38, 39, and 42. A similar polymorphism was also observed in a disrupted form of the ODV envelope protein, odv-e66 (ORF114), which has a 327-bp indel in AgMNPV-42. Moreover, the 3’ region of pe-38 (ORF006) showed a remarkable pattern of gain and loss of genomic fragments that gave rise to at least four allelic variants (PE38-α, -β, -γ, and -δ). PE38-α is the variant found in AgMNPV-2D; PE38-δ is found in CDEFMNPV, and the other three variants—β, γ, and δ—were observed, respectively, in (β) 26, 30, 32, 36, 37, 38; (γ) 27, 28, 29, 31, 34, 35, 40, 42 and 43; and (δ) 33, 39. The large variability observed in PE-38 was caused by the absence or duplication of Leucine Zipper motifs (Krapa and Knebel-Morsdorf 1991), which are structures composed mainly by leucine residues linked by glutamic acid, lysine, asparagine, and arginine (LEEKNRL), although other arrangements were also observed (fig. 3).

Gene Fusion and Fission Events

In different isolates, genes pe-38, he65, ag144, and bro-c were observed as single ORFs or divided into two independent ORFs (supplementary fig. S2, Supplementary Material online). pe-38, he65, and ag144 had their coding sequences modified due to a 1-bp indel, a single nucleotide substitution, and a 10-bp indel, respectively. The gene pe-38 is originally split into two ORFs in AgMNPV-2D. The single ORF of he65 was encoded by most of the isolates, except AgMNPV-2D, -39, and -43. In these three isolates, the N and the C terminal proteins were encoded by he65-like-1 and he65-like-2, respectively. The same pattern was also observed for ag144 that was split into two short ORFs (ag145 and ag146) in most isolates, with the exception of AgMNPV-2D, -36, and -37. Finally, in bro-c, ORFs 157 and 158 were generated in at least two independent events of gene fission/fusion. One was probably mediated by slipped strand mispairing causing a 25-bp indel (isolate 32); and the second event derived from a 1-bp indel (isolates 35 and 36). The directionality of these ORF rearrangement events could not be precisely determined due to a lack of phylogenetic signal among the aforementioned haplotypes.

Other Structural Variations in bro Genes

Proteins encoded by bro genes have high levels of plasticity and variability (Bideshi et al. 2003), and besides the structural variations involving bro-c, important changes affected two other bro genes. Alleles of bro are observed in multiple copies, varying from 1 to 16 copies (Ayres et al. 1994; Kuzio et al. 1999), although some baculoviruses lack these genes (Ahrens et al. 1997; Hyink et al. 2002). Among group I Alphabaculoviruses, isolate AgMNPV-2D has the higher number of bro genes: Seven copies (from bro-a to bro-g) (Oliveira et al. 2006). These seven copies of bro encode proteins with domains Bro-N (PF02498), DUF3627 (PF12299), and T5orf172 (PF10544), which can be covalently linked or separated in independent peptides. Interestingly, only bro-a (ORF163) is able to encode a protein with the T5orf172 domain. Interestingly, bro-a (ORF163) was absent in most AgMNPV genomes, excluding AgMNPV-2D, -26, -31, -35, -40, and -43. The mechanisms that could explain bro-a (ORF163) absence are unknown and, based on sequence information, it was not possible to conclude whether this gene was lost or acquired during the recent evolutionary history of AgMNPV. Interestingly, bro-a and bro-b were contiguous in AgMNPV, and they encode proteins with contrasting domain architectures: Bro-N + T5orf172, and Bro-N + DUF3627, respectively. However, despite their divergent architecture, a fragment of 150 bp seems to be shared by these genes, and this observation gives us clues about the origin or loss of bro-a (fig. 4). Another important variation involving bro genes took place at the locus of bro-b. This gene showed at least three variants, which could be distinguished by the absence or presence of specific codons (from 1 to 7 in total), specifically located within their interdomain region, or inside the DUF3627 domain.

Further Genetic Novelties

Besides the unique ORFs 098 (ag98) and 132 (parp-like) observed in AgMNPV-2D (Oliveira et al. 2006), four other unique ORFs were found among the wild isolates: ORFs 008 (ag8), 012 (ag12), 135 (ag135), and 139 (rnf12-like). These ORFs are not present in all isolates, as indels disrupted their coding sequences in some genomes. Both ag8 and ag12 have small ORFs, and they are found inside the most variable region of AgMNPV genomes: The locus flanked by hr11 and hr2, where bro-a, pe-38, and ie-2 are located. Ag8 was present in nine isolates (27, 28, 29, 31, 34, 35, 40, 42, and 43), whereas ag12 was observed only in AgMNPV-33 and -42.

The third new gene, rnf12-like (ORF139), potentially encodes a protein with a conserved RING-H2 Finger domain in...
its C-terminus. This domain is characterized by the motif C3H2C3, with a Hystidine replacing a Cysteine at the fourth position (Saurin et al. 1996), differing from other RING-Finger proteins of AgMNPV, which had a RING-HC Finger domain (C3HC4). A BLAST (Basic Local Alignment Search Tool) search revealed a significant match (e value = 8e-15) between this sequence and an RNF12-like protein from a lepidopteran (Bombyx mori), indicating a possible gene acquisition through an event of HGTs (fig. 5). The whole ORF of rnf12-like was found only in AgMNPV-34 and -37. In AgMNPV-2D and -32, rnf12-like was completely absent. All other isolates had a partial version encoding a small fragment of RING-H2 Finger (fig. 5). Seemingly, in this isolate this coding sequence was replaced and disrupted by an exogenous sequence composed mainly of a GT-dinucleotide repeat and a direct repeat (dr13; 4 × 15 bp).

Finally, a 45-aa polypeptide encoded by a nested ORF internal to fgf (ORF134), detected in the AgMNPV-2D proteome (Braconi et al. 2014), was found to be conserved in all AgMNPV genomes studied. As ORF135 (ag135) is located in opposite orientation to the fgf sequence, the protein it encodes does not match any region of FGF. The protein of this hypothetical gene is carried inside AgMNPV virions.

Genetic Diversity of AgMNPV Genes

The genetic diversity was calculated using both the Watterson’s genetic diversity estimator ($\theta$) and the number of nonsynonymous changes ($d_N$), which were obtained from multiple sequence alignments (MSA) of whole genomes and of 151 genes shared by all AgMNPV. The gene MSA was used to estimate the diversity as a function of the number of segregating sites ($\pi$), where all genomes (including AgMNPV-2D) were assumed as genotypes of a hypothetical metapopulation of AgMNPV. Values of $\pi$ were divided into four categories, from A to D (in intervals of 0.01), where category A, in turn, was divided into four subcategories, from A1 to A4 (in intervals of 2.5 × 10^{-3}) (fig. 6). The whole-genome MSA was used to detect nonsynonymous changes, and the total of polymorphisms per gene was plotted as absolute numbers in figure 6. Overall, at least 714 nonsynonymous changes were found, which are not evenly distributed along the genomes. Among genes shared by all AgMNPV, 109 genes grouped within the subcategories A1 and A2, as they had low $\theta$ diversity, and most of them had two or less nonsynonymous changes. A total of 79 genes were observed within the subcategory A1, of which nine were 100% identical in all genomes, they were ac29-like, ac145-like, ac110-like, bro-a,
broker, ctl-2, emt, iap3, and odv-e18. Contrasting, only 17 genes were included within the other three categories (B–D). These genes carry more than 40% of all amino acids substitutions, and they are mainly observed in a highly variable region located between hr11 and hr2 (from 12 o’clock to 1 PM in the physical map) (fig. 6). A total of 12 genes with high diversity were located within or flanking this region: ie-2, ag10, pe-38, pk1, ag4, lef-2, ac4-like, bro-b, ptp-2, ac11-like, bro-c, and anf-1. These 12 genes accounted for 306 nonsynonymous changes (42% of the total).

Based on $\theta$ values of core and noncore (i.e., satellite) genes, the level of sequence variation between these two groups of genes was evaluated. Core genes constituted a set of 37 coding sequences shared by all baculoviruses, and in AgMNPV they were observed to have significantly low diversity ($\theta = 0.0031 \pm 0.0028$) when compared with noncore genes ($\theta = 0.0067 \pm 0.0118$) (fig. 7A). In terms of amino acid changes, the same pattern was observed, given that the average number of nonsynonymous changes in core genes was smaller than in noncore genes (table 3). Interestingly, from 19 core genes with very low genetic diversity ($\theta$ subcategory A1), 10 encoded structural proteins associated with BVs or ODVs (Braconi et al. 2014): 38k, p40lc42, vp1054, vp91/p95, p6.9, odv-e18, odv-e27, odv-ec43, pif-1, and pif-3. In general, core and satellite genes encoding structural proteins had the lowest average number of amino acid changes (3.14 ± 3.66). Some of these proteins had no changes, such as those encoded by p24, vp1054, odv-e18, pif-3, and polh (table 3). Interestingly, when protein size was plotted against number of nonsynonymous changes, we have found just a trend for positive correlation between these variables ($r = 0.47$), as some genes fell distant from the trend line (fig. 7B). These results gave us an interesting overview of the distribution of proteins in terms of polymorphisms along amino acid sites, where outliers, identified as those genes not included in the 95% CI in figure 7B, were either short proteins with high number of polymorphisms, or vice versa. Among these proteins, ten were short proteins (around 200–400 amino acids in length) that had a high number of changes (from 15 to 70). These proteins were encoded by ac11-like, ac91-like, bro-c, bro-b, ac34-like, pe-38-like-1, pk1, and ie-2 (fig. 7B). The last four genes in this list are known to be involved in DNA replication and/or transcription. This pattern was also observed for dnapol, which encodes the second largest protein of AgMNPV (991 aa), and was highly variable (22 changes). Conversely, not all long proteins appeared to have accumulated a high number of nonsynonymous changes. A high level of sequence conservation was observed in desmosplakin (886 aa) and helicase (1,221 aa), which had only two and six changes, respectively (fig. 7B).

**Discussion**

**The Genome Compaction of AgMNPV**

As shown in figure 2, most of the AgMNPV ORFs have short or absent intergenic spacers, and this high level of genome compaction directly affects how viral genomes evolve (Chirico et al. 2010). Even cis-acting elements such as promoters, transcription starting sites, and polyadenylation sites can lie within adjacent coding sequences (Normark et al. 1983; Chen et al. 2013). Because mutations in these shared regions can impair more than one genetic element (whether coding or regulatory), these overlaps impose constraints on sequence evolution (Krauker 2000). Moreover, as breaking points are likely to be deleterious when within overlaps, this high level of genome compaction could explain why genome collinearity was conserved among AgMNPV genotypes, as recently observed in other Alphabaculoviruses (Theze et al. 2014; Chatteigner et al. 2015).

**Polymorphisms in hrs**

In previous studies involving other genotypes of AgMNPV, hrs were shown to be highly variable, differing from each other on their number of tandem repeats and core palindromes (Garcia-Maruniak et al. 1996; Maruniak et al. 1999). Among the 17 wild-type genotypes analyzed in this work, these copy-number variations were specially observed in large hrs, such as...
Remarkably, the haplotypes flanked by these hrs were the most variable in terms of gene content. Between hr11 and hr3 are variable genes as pe-38, 1629capsid, new genes as ag8 and ag12, as well as bro-a (absent in some isolates). On the other hand, the haplotype flanked by hr9 and hr10 encompassed the polymorphic gene hisP, and the parp-like (described previously in AgMNPV-2D) and mfl2-like (found as partial or full-length variants in some isolates) genes that are specific to the AgMNPV. Based on the repetitive nature of the hrs, and taking into account their roles as origins of replication (Habib and Hasnain 1997), the size variation observed within hrs could be the result of slipped strand mispairing during events of DNA replication (Levinson and Gutman 1987; Bzymek and Lovett 2001), and the genetic
these repetitive elements (Gemayel et al. 2010), as hrs work as enhancers for the expression of early genes (Olson et al. 2001; Landais et al. 2006). Size variations in hrs are common among alphabaculoviruses, such as SfMNPV (Simon et al. 2011), AcMNPV (Chateigner et al. 2015), and MacoNPV (Li et al. 2005), in which extensive indels were found.

**The Dynamics of Gain and Loss of Bro Genes**

In the AgMNPV, only bro-a (ORF163) encodes a polypeptide containing a T5orf172 domain. This gene is usually located adjacent to bro-b (ORF162), and flanked by ptp-1 and ac4-like (Oliveira et al. 2006). Gene content in this region is highly variable in different baculoviruses, especially due to partial or complete deletion of homologues of bro-a (Supplementary fig. S3, Supplementary Material online). In AgMNPV, 12 isolates lacked bro-a, showing a haplotype similar to those observed in AcMNPV (Ayres et al. 1994) and CidDefMNPV (Lauzon et al. 2005). Moreover, the partial or complete lack of bro-a was recently suggested to be associated with multiple events of deletion in BmNPV isolates (Ardisson-Araujo et al. 2014). Both the structural variations and the similarity observed between bro-a and bro-b in AgMNPV (fig. 4), allowed us to hypothesize on the evolution of the bro-a/bro-b haplotype. Possibly the Bro-N domain encoded by bro-a could have been derived from a partial duplication of bro-b, followed by an independent acquisition of the T5orf172. Meanwhile, bro-a and bro-b could have been acquired together from a viral or a prokaryotic genome (Iyer et al. 2002). Moreover, the deletion of bro-a/b could have taken place during genome replication, whether due to intramolecular slipped-strand mispairing between bro-a and bro-b, or due to asymmetric homologous recombination between related genotypes (for details, see supplementary fig. 3, Supplementary Material online). In AgMNPV, the locus between hr11 and hr2 gathers at least one-third of the nonsynonymous changes that we detected. As shown in figure 6, this region encodes three bro genes: bro-a, -b, and -c. The latter two genes accumulated more than 11% of the amino acid changes reported for AgMNPV. Interestingly, this pattern was also observed in MacoNPV-A, in which gene gains/losses and high genetic diversity are observed in bro genes located specifically within a highly variable region (Li et al. 2005).

**Gene Fusions/Fissions and Size Variations in Genes of AgMNPV**

We presented at least four examples of gene fission and fusion in AgMNPV genes, events driven by point mutations (1-bp indels/substitutions) and short indels (10–25 bp). As observed in AcMNPV, the fusion of adjacent genes is a common mechanism of evolution among alphabaculoviruses (Chateigner et al. 2015). In the AcMNPV, at least four pairs of adjacent ORFs were found to be fused (Ac20/Ac21, Ac58/Ac59, Ac106/Ac107, and Ac112/Ac113), but these pairs do not correspond to those observed in our results.
Table 3
Statistics of Nonsynonymous Changes (NSC) among Genes from Different Functional Categories

<table>
<thead>
<tr>
<th>Functional Category</th>
<th>Average NSC and Standard Deviation</th>
<th>Genes and Number of NSC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Auxiliary</td>
<td>5.54 ± 11.01</td>
<td>25k fp (2); ac145-like (0); ac4-like (4); ac5-like (2); ag127 (1); ag128 (2); ag22 (1); bro-b (49); bro-c (34); bro-d (3); bro-e (0); bro-f (3); bro-g (0); cg30 (4); ctf-2 (0); desmplakin (2); gp16 (0); he65 (5); hisP (6); ac141-like/exon0 (2); p48 (1); parp-like (3); pkp31 (6); pp31 (6); sod (3); v-trex (2)</td>
</tr>
<tr>
<td>Host modulation</td>
<td>4.00 ± 4.24</td>
<td>arf1-1 (13); egt3 (3); tfg (0); iap-1 (2); iap-2 (2); iap-3 (0); p12 (0); p33 (6); ptp-1 (5); ptp-2 (10); v-ubi (3)</td>
</tr>
<tr>
<td>Replication</td>
<td>7.73 ± 7.50</td>
<td>alk-exo (6); dbp (3); dnapol (22); gta (4); helicase (6); lef-1 (2); lef-2 (9); lef-3 (1); lef-7 (2); me53 (6); pe-38-like-1 (22)</td>
</tr>
<tr>
<td>Transcription</td>
<td>8.56 ± 17.47</td>
<td>39k/pp31 (2); ac38-like (2); ie-1 (3); ie-2 (70); lef-10 (1); lef-11 (0); lef-12 (1); lef-4 (5); lef-5 (6); lef-6 (1); lef-8 (5); lef-9 (6); met (5); p47 (3); pk1 (26); vf1 (1)</td>
</tr>
<tr>
<td>Capsid</td>
<td>3.83 ± 5.25</td>
<td>1629capsid (18); 38k (1); ac53-like (1); p15 (1); p24 (0); p25 (3); p40 (3); p6.9 (1); p87 (9); vp1054 (0); vp39 capsid (2); vp91/p95 (7)</td>
</tr>
<tr>
<td>Virion</td>
<td>2.78 ± 2.56</td>
<td>ac108-like (1); ac150-like (1); ac68-like (1); ac81-like (2); f-protein (10); gp41 (2); gp64 (4); odv-e18 (0); odv-e25 (3); odv-e26 (5); odv-e27 (2); odv-e56 (7); odv-e66 (5); odv-ec43 (2); p10 (2); p49 (1); p74 (7); pif-1 (4); pif-2 (2); pif-3 (0); pif-4 (2); polh (0); pp34/pep (1)</td>
</tr>
<tr>
<td>Core genes</td>
<td>3.65 ± 3.97</td>
<td>See genes in bold above.</td>
</tr>
<tr>
<td>Noncore genes</td>
<td>4.91 ± 9.27</td>
<td>—</td>
</tr>
<tr>
<td>Total</td>
<td>4.61 ± 8.32</td>
<td>—</td>
</tr>
</tbody>
</table>

Note.—Genes highlighted in bold correspond to core genes. Functional categories were assigned based on Cohen et al. (2009).

(indicated that the conservation of promoters determines how these ORF rearrangements impact the viral proteome, and consequently, on baculovirus infection.)

A Nested Gene in AgMNPV

Small reading frames nested within large ORFs are common in viral, bacterial, and mitochondrial genomes (Normark et al. 1983; Krakauer 2000). Using mass spectrometry data recently obtained by our group, a 45-aa peptide was found to be encoded by an ORF located internally and in opposite orientation to tfg (ORF134). The short peptide encoded by this nested gene (ag135) seems to be expressed and carried along with AgMNPV virions (Braconi et al. 2014), and this might be the first report of a protein encoded by a nested gene in baculovirus. However, further studies are necessary to characterize the promoter motif, gene expression, and function of this new protein in AgMNPV.

Acquisition of an E3 ubiquitin-Ligase Gene

A new gene, mf12-like (ORF139), was observed in at least two wild populations of AgMNPV (34 and 37), and it encodes for a protein containing a RING-H2 Finger domain on its C-terminus. RING Finger domains are also encoded by other AgMNPV genes, such as cg30 (Passarelli and Miller 1994); the antiapoptotic genes, iap-1, iap-2, iap-3 (Clem and Miller 1994), ie-2 (Passarelli and Miller 1993), and pe-38 (Krappa et al. 1995), but their RING-HC Finger domains differ from the one encoded by mf12-like (fig. 5). Proteins with any of these domains are known to belong to different functional categories (Freemont 2000). However, those with RING-H2 Finger are specially involved in the ubiquitin pathway, acting
as E3 ubiquitin-protein ligases (Joazeiro and Weissman 2000). Among baculoviruses, three proteins with RING-HC Finger have already been pointed as E3 ubiquitin-protein ligases: IAP2, IE2, and PE38 (Imai et al. 2003). Almost all Alpha- and Betabaculoviruses express an additional component of the ubiquitin pathway: The viral ubiquitin, encoded by v-ubi (Guarino 1990). In AgMNPV, the gene v-ubi (ORF138) is located upstream of mf12-like (ORF139), in the same orientation. As the ubiquitin–proteasome system has a central role targeting proteins for degradation (Katsuma et al. 2011), we hypothesize that the physical proximity of these viral genes could somehow be advantageous for their coexpression.

Based on the pattern of indels within the mf12-like locus (fig. 5), the AgMNPV has probably acquired this gene through HGT, and then has gradually lost parts of this coding sequence in at least two moments over its evolutionary history. It is well known that the genomes of baculoviruses and other large DNA viruses evolve through subsequent events of HGT involving eukaryotic hosts, recombination with other viruses, and gene duplication followed by sequence divergence (Hughes and Friedman 2003; Shackelton and Holmes 2004; Zanotto and Krakauer 2008).

Most AgMNPV Genes Have Low Sequence Diversity

To analyze the genetic diversity in AgMNPV, we used two different metrics—θ and number of nonsynonymous (i.e., amino acid) changes. It is important to highlight that both measures of genetic diversity rely on nucleotide substitutions, but their results are slightly different, and are not necessarily proportional to each other. As these analyses were performed based on consensus genome sequences reconstructed from heterogeneous field isolates, a potential limitation of our results would be the assumption of clonality in AgMNPV populations, phenomenon uncommon for natural NPV isolates. Nevertheless, we preferred to summarize each population by its most frequent, master consensus sequence for expediency, which we assumed to be the predominant genotype of each viral population. Notwithstanding, our analyses revealed that 99 out of 167 AgMNPV genes had low sequence diversity (two or less polymorphisms), and core genes are on average less diverse than satellite genes, reinforcing the observations made by Oliveira et al. (2013). A total of 25 out of those 99 conserved genes encode structural proteins associated with viral envelope, nucleocapsid, or nucleosome formation (Bracconi et al. 2014), and interactions among them or among these proteins with DNA are common (Rohrmann 1992). As changes at the sequence level can impair such interactions, we might expect that these proteins would be evolving slowly under purifying selection (Daugherty and Malik 2012), although other hypotheses to explain this observation are not discarded.

At the other extreme of diversity, a particular genomic region delimited by hr11 and hr2, approximately 18,000 bp grouped several genes with high sequence diversity (fig. 6). An increase in variability, such as the one around 12 o’clock in the physical map of the AgMNPV pangenome (fig. 6), was also observed in two isolates of MacoNPV, which have that region accumulating at least 50% of the total nucleotide changes, most of them located in two bro genes (Li et al. 2005). Nevertheless, this variable locus of MacoNPV does not match the one AgMNPV, as they differ in gene content. Interestingly, the variable locus at 12 o’clock in AgMNPV was shown to evolve through homologous recombinations, as reported in wild isolates studied by Crozier and Ribeiro (1992). In that work, different restriction sites were used as genomic markers to trace recombination between AgMNPV genotypes. One of these markers (named “e”) is located exactly in the locus of ie-2, here presented as the most divergent gene, showing at least 70 polymorphisms. Curiously, this locus gathers several genes whose products act as trans-acting factors, namely: acd-like, ac5-like (Lo et al. 2002), ac141-like/exon0, ie-1 (Kool et al. 1994), ie-2 (Prikhod’ko et al. 1999), and pe-38 (Krappa et al. 1995). This variable region also contains three bro genes (bro-a, -b, and -c), all of them encoding the domain Bro-N. This domain has been shown to have DNA-binding activity, and probably acts as transcriptional regulator of viral and host genes (Zemskov et al. 2000). The high level of polymorphism and the cluster of transcriptional regulators encoded there suggested that the variable region flanked by hr11 and hr2 may play a relevant role in determining the infectivity of AgMNPV. As this region frequently evolves through homologous recombination, it could facilitate gene exchange among AgMNPV genomes. Overall, the presence of variable regions and polymorphic alleles in baculoviral populations are important evolutionary advantages, once it is shown that these variable elements enable these viruses to respond rapidly to selective pressures (Li et al. 2005; Simon et al. 2011; Chatteignier et al. 2015).

Conclusion

In this study, the complete genomic sequences of 17 wild isolates of Anticarsia gemmatalis MNPV were sequenced and characterized. A total of 167 predicted genes were observed in the pangenome of the AgMNPV. Among them, 151 are shared by all isolates, including the reference isolate 2D; six were implicated in events of gene fusion/fission; and other four were only found in a few genomes, including bro-a, and the new gene mf12-like, which encodes for an E3 ubiquitin-ligase probably acquired through HGT. Gain and loss of small genome fragments were mainly observed in tandem repeats located inside or outside coding sequences. Most AgMNPV genes had low nucleotide diversity, and variable genes were mainly observed in a region previously described to be involved in homologous recombination. This study provides relevant information about AgMNPV genome evolution in particular and on baculovirus in general.
Supplementary Material

Supplementary figures S1–S3 and tables S1 and S2 are available at Genome Biology and Evolution online (http://www.gbe.oxfordjournals.org).

Acknowledgments

A.F.B. and P.M.A.Z. thank Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) and Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP, 2011/14537-0 and 2011/17120-3) for the financial support; C.T.B. holds a postdoctoral fellowship (FAPESP 2014/03911-7). A.G. and P.M.A.Z. hold a CNPq/PQ Scholarship. The authors declare that they have no competing interests. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Literature Cited


Associate editor: Tal Dagan.